

White Paper – Combining Attitudinal Data and Behavioral Data for Meaningful Analysis

By Abigail Lefkowitz, MaassMedia

Executive Summary:

In the fast-growing digital analytics industry, it is important for analysts to provide their clients with actionable insights to uncover overlooked opportunities for growth and to highlight neglected areas that require immediate attention or long-term improvement.

Customer satisfaction (C-SAT) surveys aim to investigate these aspects of the user experience, but I have found that most surveys fall short of presenting a multi-faceted view of the data. C-SAT scores do not always show much change from month to month, making it tricky to produce a meaningful analysis from satisfaction data alone. Often, a second set of data is necessary to layer upon customer satisfaction data in order to expose deeper meaning. Data from web analytics tools can provide that layer.

In this article, I will discuss:

- The challenges faced when attempting to find meaning in C-SAT data
- How to combine C-SAT data with web analytics data to gain more useful insights
- The most appropriate types of analyses to conduct

The Problem:

What is Customer Satisfaction?

A customer satisfaction score is a measure of how well a company is satisfying its customers. This score is often a calculation based on a number of factors and is compiled from surveying customers either on- or offline.

C-SAT scores can be compared across companies within the same industry using the standard scoring system set by the American Customer Satisfaction Index (ACSI). The score is a number ranging from 1-100 and is calculated based on a set of scaled survey questions. A lower score indicates customer dissatisfaction, while a higher score implies customers are generally

satisfied. While this ACSI score is easily comparable on a nationwide basis, most companies also have internal reports of C-SAT that are derived from different metrics of their preference.

C-SAT scores can provide meaningful information, especially when the numeric score fluctuates substantially over reporting periods. But what can be learned from a C-SAT score that remains relatively constant? While some may think a consistently high C-SAT score is a good thing, little useful information can be gleaned from a value that does not change. In order to get more out of a C-SAT score, other behavioral data must be factored in.

The Nature of the Data

Merging attitudinal survey data and behavioral web data has many advantages over analyzing either one individually. Attitudinal surveys measure only what respondents self-report. You might assume that people are, not only honest, but also unbiased in their responses, but this is largely not the case. Many outside elements such as time of day, personal problems, financial struggles, or even weather can change the way someone might respond to a survey. Therefore, survey data is inherently subjective.

While attitudinal data only captures information about what people say they do, behavioral data focuses on what people actually do. Because of this, behavioral data is always objective and free of bias. The downside is that it offers no insight into why a person might have viewed a certain string of pages, searched for a specific term, or watched a number of videos.

Why Do We Need Both Attitudinal and Behavioral Data?

Combining both subjective and objective data allows analysts to paint a more comprehensive picture of a visitor's experience as a whole. Be advised that the population from which the sample is drawn must be specifically and carefully defined. The sample reflects only those who answer a survey, so it would be inaccurate to say it represents the entire population of visitors. When measuring engagement, it should also be noted that visitors who fill out surveys are generally more engaged with the site's content than visitors who do not fill out surveys.

Our Solution:

The purpose of this solution is to:

- Provide a methodology for acquiring, merging, storing, and analyzing attitudinal and behavioral data
- Enhance insights and bring depth to superficial customer satisfaction measures

Acquiring the Data

Assuming customer satisfaction data is readily available through the chosen survey interface, the first step is to gather the behavioral data. Web analytics code from a tool such as Google

Analytics or SiteCatalyst must be implemented on the site in order to obtain the appropriate metrics. Both tools provide access to the data in their own user interfaces and give several options for exporting the data. Make sure to capture the unique survey identifier as a metric to allow for easy merging with the survey data.

Because response rates to surveys are generally low (3-5%), several months or years of data is necessary to ensure a decent sample size for valid analysis. Once the appropriate data has been exported into a usable format, any number of data warehouse programs can be used to merge and store the data. This solution utilizes Microsoft Access. Merge the data by unique survey identifier and remove duplicates and outliers when appropriate. Load the merged data into a preferred business intelligence (BI) analysis tool and begin the analysis.

Assumptions and Hypotheses

It is always prudent to outline the assumptions and formulate hypotheses before delving into any analysis. A detailed list of assumptions including time frame, variables included, metric definitions, and all calculations helps define the clarity and scope of the project. Statistical tests require a separate set of assumptions and should be listed where appropriate.

Having several testable hypotheses gives the project a purpose and reduces the time spent poking around in the data with no clear direction. That is not to discount aimlessly wandering through the data, because some of my best insights have happened by accident; it is merely a best practice for adhering to traditional research methodology. In general, I spend about 75% of the time investigating hypotheses and 25% stumbling upon unanticipated findings.

When forming hypotheses, there are many questions to ask about the data. For example:

- Does visitor satisfaction change after a web site or page redesign?
- What are the biggest drivers of satisfaction?
- What are the most common barriers to raising satisfaction?

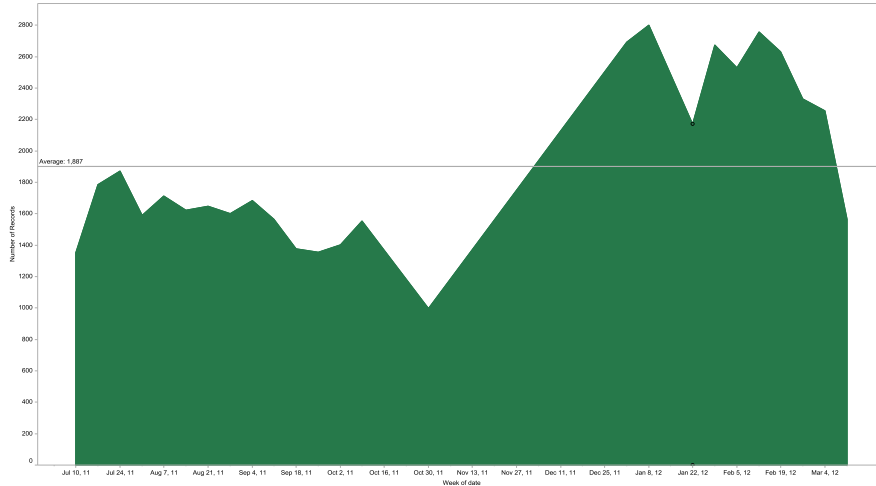
Preliminary Analysis

Any number of statistical analyses can be performed on the data, ranging from elementary to advanced; however, high-level testing is rarely appropriate or significant. Traditional and more basic methods including descriptive statistics, linear regression, and goodness of fit testing are the most useful for discovering deeper insights into the data.

Preliminary analysis should focus on describing the data in its rawest form. Plotting number of records over time can provide a general idea of how responses vary over week, month, and year. Are there seasonal trends? Was there a specific event that caused a spike or drop in responses? Matching this data to number of visits to the site will give greater insight into

whether or not a drastic spike in responses is due to a corresponding spike in web traffic for the same time period.

The area chart shows number of responses over time by week. There is a sharp decline the week of October 30, 2011 followed by a steep increase through the week of January 8, 2012. The obvious assumption is an increase in traffic due to the holiday season, but without syncing this chart to site visits, such a conclusion cannot be soundly validated.



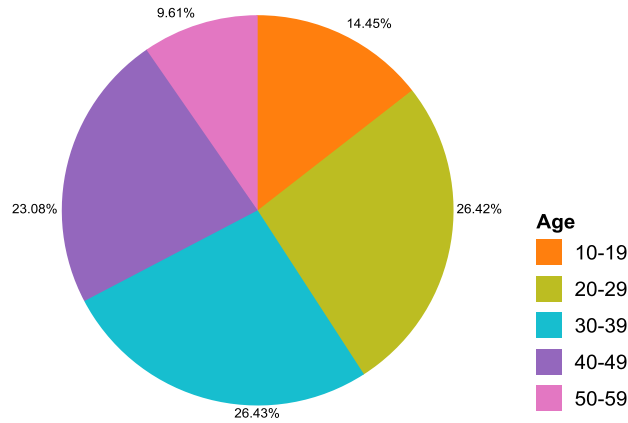
Describing the Data with Demographics

Demographics such as gender, age, marital status, income level, and geographic location can be segmented to provide a snapshot of what the sample looks like. Some questions to investigate:

- What age groups respond most frequently?
- Is there a region where respondents are more concentrated?
- Does number of responses differ by gender?

While these findings may not prove the most useful, they are necessary to understand the sample.

The pie chart below shows percent of respondents segmented by age group. It is helpful to see if one slice is considerably smaller than the other slices in the chart because its contribution to the data will not be as significant as a larger slice and should be noted as such.

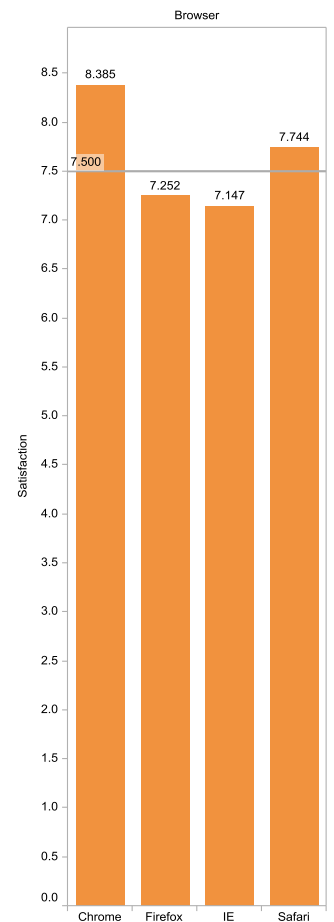


Comparison to Average

Preliminary analysis should also establish some sort of benchmark or average value to which segmented data can be compared. Depending on the purpose of the analysis, this number can be anything from a measure of satisfaction (C-SAT), number of pageviews, or even revenue. Using this average as a baseline for comparison will reveal any significant differences among the data that might not have been noticeable on their own.

In the bar chart to the right, the reference line represents average customer satisfaction for the time period analyzed. Plotting this average on the chart reveals whether satisfaction differs when segmented by browser. It is clear that those using Chrome report higher average satisfaction than those using Internet Explorer, but without the reference line one would not know how these compare to the overall average.

Statistical testing for significant differences between average satisfaction can be carried out from this point; however, it may be difficult to find statistical significance among the data when the sample size is considerably large and the range of values tends to linger around the same number. ANOVA testing for a difference between multiple means, the appropriate test, will rarely generate significant results and should be avoided.

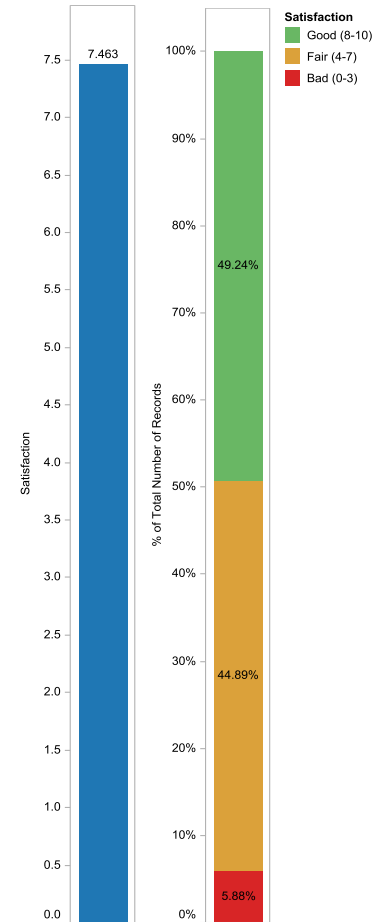


Visualizing Average Satisfaction

Average satisfaction can be represented either by overall average or by "box" rating. Visually, overall average is simply a bar with height corresponding to the specific number. Average by "box" shows the distribution of responses as percentage of the total number of records.

Depending on how the possible responses are defined, there are can generally be anywhere from two (0 or 1) to eleven boxes (0-10).

The bar chart to the right shows both representations of satisfaction, with average on the left and percentage of total by box on the right. Each tells a different part of the story. The raw number represents satisfaction, while percentage of total by box shows a more detailed breakdown of the responses that contribute to the raw number. In this case, values from 0-10 were grouped into three categories for simplicity. The groups can certainly be expanded or condensed for more or less granularity. Keep in mind the audience to whom the data is being presented; those with an advanced background in analytics are more inclined to favor detailed graphs with multiple axes and several charts per pane, while less analytical clients may prefer simpler visuals with clearer guidance.



Analysis by box is useful for examining the "top two" and "bottom two" boxes of customer satisfaction. In general, there are three main box classifications:

- Top two - responses including "good," "very good," "above average," "outstanding" (green)
- Bottom two - responses such as "very bad," "bad," "below average," (red)
- Middle - anything along the lines of "fair," "okay," or "average" (yellow)

Ideally, companies want to decrease the number of responses in the bottom two boxes, while simultaneously increasing the number of responses in the top two boxes. This would result in a higher overall satisfaction raw number. Usually, this happens slowly over time with dynamic improvements to the customer experience.

Depending on the nature of the data and the goal of the client, analysis can include other statistical testing such as chi-square goodness of fit, linear regression, and factor analysis.

Solution Benefits:

In order to deliver a multi-dimensional analysis, reports of customer satisfaction must include behavioral metrics gathered outside of the survey. Merging attitudinal and behavioral data allows further exploration of C-SAT data beyond the obvious conclusions.

Clients will be able to use the resulting analysis to:

- Improve the user website experience by identifying easy fixes and long-term changes that must be made
 - Optimize site usage by removing underutilized tools, combining similar site sections, and streamlining content
 - Raise overall conversions by focusing users on completing essential goals like purchasing, signing up for newsletters, or taking advantage of deals and offers
- Multidimensional analytics, the combination of data from multiple sources, help break down information silos, offer richness, and if correctly executed and interpreted, facilitate deeper insights about consumer behavior. However, these analytics are not easy to accomplish. Having these capabilities distinguishes the "great" from the "good."

Appendix:



Technical Tools

Web Analytics Tools

- Google Analytics
- Site Catalyst

Business Analytics/Intelligence Tools

- Tableau
- R
- SAS
- SPSS
- Minitab

<p>Certifications:</p> <div style="display: flex; align-items: center;">  <div style="text-align: center;"> <p>ANALYTICS QUALIFIED INDIVIDUAL</p>  </div> </div>	<p>Contact Info: Abigail Lefkowitz MaassMedia, LLC 1315 Walnut St., Suite 1108 Philadelphia, PA 19107 (267) 296-9657</p>
---	--