Columbia Business School
AT THE VERY CENTER OF BUSINESS™

Columbia CaseWorks

Web Analytics at Quality Alloys, Inc.

BY ROB WEITZ[*] AND DAVID ROSENTHAL[†]

Company Background

Quality Alloys, Inc. (QA)[1] is a relatively small (less than $75 million in annual sales) US-based distributor of different grades of a variety of alloys used in industrial manufacturing. Its products include, for example, refractory alloys, which have the property of maintaining strength at high temperatures and are used in furnaces and aircraft turbine engines.

QA's market niche is based on a number of factors:

1. It has expert knowledge regarding which suppliers manufacture the different grades of alloys and those who do so reliability.
2. It sells in small quantities.
3. It cuts material to user-specified sizes, with tolerances of less than .001 of an inch.
4. It ships items in stock the same or next day.

QA's customers are generally small companies that make parts out of the alloys they purchase. They generally can't buy directly from mills (the alloy producers) as their orders are not large enough. These customers do not inventory the alloys purchased from QA, and they typically purchase what they need for specific jobs they have in hand. One purchase may not be an accurate predictor of future purchases; that is, one purchase may not signal what will be purchased in the future, or when a future purchase will occur. The individuals making the purchasing decisions may be engineers, purchasing agents, or the business owners.

QA is well-respected in its market space. Its competitors are other intermediaries providing similar services.

Author affiliation
[*]*Adjunct Associate Professor, Columbia Business School and Associate Professor, Stillman School of Business, Seton Hall University*
[†]*Associate Professor, Stillman School of Business, Seton Hall University*

QA markets its products through direct mail, advertising in trade magazines and, more recently, via paid listings on two industrial product web portals, GlobalSpec and ThomasNet. In mid-2008 QA decided to extend its marketing reach by establishing a company web presence. The goals of the company's website were to (a) drive new sales, (b) make product and contact information available, and (c) give or add legitimacy to its brand. Further, in addition to providing QA with another opportunity for reaching out to its traditional clients, management thought the website would enable QA to extend its reach to many more customers in the United States, Europe, and Asia. The Asian market was viewed as particularly important to QA given the shift towards manufacturing in the Pacific Rim.

Given the intricacies of QA's products, the website was not designed to allow users to enter orders over the web. However, in addition to providing descriptions of the products and services offered by QA, the website does allow potential customers to submit request for quotations; that is, a request to be contacted. These requests are given to inside sales staff for follow up. As is the case for other business-to-business (B2B) providers, there's no direct way to connect a visit with a sale as there is no "shopping cart" on the website.

QA commissioned a professionally produced brochure providing an overview of the company's products and services and sent it to potential customers in mid-December 2008. The list of potential customers and their addresses was purchased by QA. The total cost of the promotion is estimated to be approximately $25K.

There are a number of questions that would be useful for QA to have answered before investing in further promotional activities. Foremost of these are the following:

- How many people visit the website? How do they come to the website?
- Is the website generating interest, and does this interest yield actual sales?
- Do traditional promotions drive web traffic, and in turn drive incremental sales?
- How can visits to the website best be modeled?
- Where and how should QA advertise?

Business Value Assessed

While the usefulness of B2B websites in terms of advertising, sales promotion, public relations, and marketing is generally recognized, valuing B2B websites remains elusive. For business-to-consumer (B2C) websites, the valuation task is facilitated by the fact that purchases are typically made via the website. Data can be compiled on sales volume, the effect of promotions on sales, how website design changes impact sales, etc. For B2B websites without "shopping cart" functionality the task is considerably more challenging. A principal conclusion of a Forrester Report on B2B marketing is that "business-to-business (B2B) marketers collect mounds of data . . . but still struggle to find effective ways to measure and demonstrate success." Some 44% of the survey respondents cited "demonstrating ROI" as the top challenge in managing their website.[2]

More generally, there has been a longstanding debate on the value of organizational investments in information technology (IT). Some have argued that investments in IT are

closely linked to organizational effectiveness.[3] Others, on the other hand, have gone so far as to argue that "IT doesn't matter."[4]

Web Analytics

QA's website developer utilized Google Analytics (http://www.google.com/analytics/), a free web analytics tool so that the site would track, and QA management would have access to, a wide variety of web metrics. Google Analytics is implemented by putting the appropriate code, provided by Google, on each web page. The data collected is sent to Google; users, like QA managers, access it by logging in to their Google Analytics account. Typical metrics include the number of visitors to a site, the amount of time they spent on the site, the number of pages viewed, etc.

Apart from Google Analytics, QA participated in Google's AdWords program.[5] AdWords ads are the "sponsored links" that appear next to, and sometimes above, Google search results. That is, QA paid to have these ads (with links to the QA website) appear on the Google search result page when relevant search terms (specified by QA) were used.

Website Visits

Figure 1 is a plot of the number of visits to the QA website per week over the period May 25, 2008–August 29, 2009. The start date was the first full week the website was available to visitors. The end date was selected as it coincided roughly with the launch of another QA website, which had the potential of affecting the data at the primary website. The time frame includes the direct mail brochure promotion sent during December 2008.

FIGURE 1. VISITS TO THE QA WEBSITE PER WEEK.
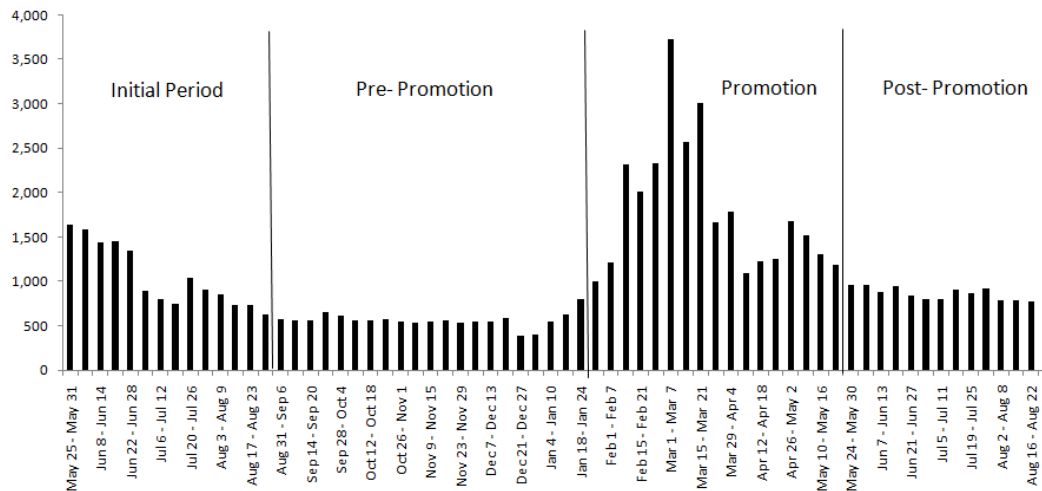
The entire time frame has been divided into four sections. The first comprises the "initial" or "break-in" period for the website. The second, established when visits appear to stabilize at a roughly constant level, is the "pre-promotion" interval. The "promotion" period begins

BY ROB WEITZ* AND DAVID ROSENTHAL†

shortly after QA's promotion took place and website visits dramatically increased. The final period, "post-promotion," begins when visits appear to level off again. Two points are worth noting. First, it seems clear from both the data and feedback from QA management that this general partitioning makes sense. For example, website data through May 2010 (not provided here) indicate that visits remained approximately at the same level as the post-promotion period; it can be reasonably assumed therefore that the jump associated with the promotion period is not due to some seasonal variation or external economic factor. Second, while there is some arbitrariness about the exact cutoffs between periods, varying them in any reasonable way does not impact the analysis.

Assignment

The Deliverable

Your job, as a trusted advisor to QA management, is to examine the value of QA's website by tracking not just website metrics (e.g., visitors, pages/visit, time per visit, etc.), but the associated (non-web-based) financial measures of sales, profit and amount of merchandise sold. One focus of your efforts should be an analysis for senior management regarding the effectiveness of QA's promotional effort.

Your report should be submitted in two parts. The first part should be an executive summary, one to two pages in length. It should summarize results and provide recommendations to QA managers as to how they might best market their business with an aim towards improving sales. Your recommendations may rely on your knowledge and experience beyond the case, but your suggestions must be supported by the data and your analyses. (You may, and probably should, do additional analyses other than those required here. You should at least review all the data.) You can use the questions from the Company Background section of this document on page three as a guide for what your summary should cover. Finally, note that this is a "real-world" case; that is, depending on your analysis you may not have all the data you want. You may certainly include as part of your recommendations that additional data be collected.

The second part of your report should be responses to the questions posed in the Analysis section that follows. Be sure your responses are numbered, corresponding with their respective questions. There is no need for added verbiage, just your answers to the questions in order. It's preferable if this is submitted in Word or .pdf format (like the first part). Excel printouts are fine as long as they are neatly formatted.

You should do the second part (the quantitative analysis) prior to doing the first part (the executive summary). **Be sure to read through the entire case at least once before proceeding.**

Analysis

DATA

The data for this case is provided in the Quality Alloys Data spreadsheet. The data is collected over weekly intervals for the period May 25, 2008–August 29, 2009, unless otherwise indicated. The number of visits per week, as well as some visit-related information captured by Google Analytics, is in the Weekly Visits worksheet. (The meaning of each variable is provided in the glossary at the end of this document.) Financial data, collected via the Enterprise Resource Planning (ERP) system at QA is provided in the Financials worksheet. (The revenue and profit variables are self-explanatory; the pounds sold variable represents the total number of pounds of material sold that week, and the inquiries variable captures the total number of inquiries received by the sales staff at QA. The records do not indicate which of these inquiries were generated directly from the website.) The Lbs. Sold worksheet contains pounds of material sold data over a broader time span, from January 2005–mid-July 2010. (This is the only data collected over a time frame other than May 25, 2008–August 29, 2009.) The Daily Visits worksheet contains, as you might expect, the number of daily visits to the QA website. Various "demographic" data for site visitors, also from Google Analytics, is provided in the Demographics worksheet.

DESCRIPTIVE STATISTICS

Start by getting a better sense of the data in each of the four periods.

1) Using data in the Weekly Visits and Financials worksheets, create four column charts (like Figure 1: Visits to the QA Website per Week) for unique visits over time, revenue over time, profit over time, and pounds sold over time. You do not have to indicate on these charts the cutoffs for the four periods.

2) Using the same data, calculate the following summary statistics for visits, unique visits, revenue, profit, and pounds sold: mean, median, standard deviation, minimum, and maximum, for the initial, pre-promotion, promotion, and post-promotion periods. So, for each period you should provide 25 values: five summary measures for each of five variables, as per the table below for the initial period.

VISIT AND FINANCIAL SUMMARY MEASURES–INITIAL PERIOD

	Visits	Unique Visits	Revenue	Profit	Lbs. Sold
mean					
median					
std. dev.					
minimum					
maximum					

That is, you should create four such tables, one for each period.

Suggestion: It's probably easiest to start by copying the data for these five variables to a new spreadsheet. Then either:

   a) Use the Descriptive Statistics tool for the data for each period, and then create the four tables from this output, or

BY ROB WEITZ* AND DAVID ROSENTHAL†

Columbia
CaseWorks

b) Create a (blank) table like the one above to the right of the initial period data. Use the Excel functions to calculate the summary measures for visits for the initial period. You can then copy these values across to complete the table for the initial period. Now create another blank table for each additional period and repeat.

> **Excel Tip: Descriptive Statistics.** You can determine these summary values using the Excel functions =average(), =median(), =stdev(), =min(), and =max(). You can also get a variety of summary measures, including these five, by using the Descriptive Statistics option: click on Data/Data Analysis/Descriptive Statistics.

3) Create a column chart of the mean visits over the four periods—that is, your chart should have four columns, the first representing the mean visits during the initial period, the second representing the mean visits during the pre-promotion period, etc. Create four more such charts, this time using the mean unique visits, mean revenue, mean profit, and mean pounds sold statistics.

Suggestion: Create another (blank) table on the same worksheet that looks like the one below. Type the cell addresses for the mean visits for each period (that you already calculated) into the appropriate cells in the first column, then copy these values across to complete the table. Creating the column charts from this table is straightforward.

	Means				
	Visits	Unique Visits	Revenue	Profit	Lbs. Sold
Initial					
Pre-Promo					
Promotion					
Post-Promo					

4) Write one or two paragraphs summarizing your findings thus far. Be sure to describe the behavior of each variable. Indicate what the results seem to show about the relationships between the variables, and the apparent effect(s) of the promotion. (In the next section you'll explore this further; feel free to make any conjectures here that seem reasonable.) Be sure to support your verbiage with your analysis results.

RELATIONSHIPS BETWEEN VARIABLES

Up until now, you've analyzed each variable separately. It would seem reasonable to look at pairs of variables and see what happens to one as the other varies.

5) Start by taking a look at revenue and pounds sold. (Before proceeding, what does your intuition say about the relationship between these two variables?) Create a scatter diagram of revenue versus pounds sold. (Revenue should be on the y, or vertical, axis.) Determine the correlation coefficient of revenue and pounds sold.

> **Excel Tips: Scatterplots**. 1) You begin creating a scatter plot by highlighting two columns of data. If the columns are not contiguous, you can highlight the second column by holding the CTRL key down while selecting those cells. 2) By default Excel will place the variable in the

right-most column on the y (vertical) axis of the scatter plot. If you wish to change this, right-click on a data point in the scatter plot, choose Select Data in the pop-up menu, click the Edit button, and then edit the Series X values and Series Y values entries as appropriate. 3) Be sure to change the chart heading to "Y vs. X" where Y is the variable on the vertical axis and X is the variable on the horizontal axis, label the axes, and delete the (unnecessary) legend.

Excel Tip: Correlations. You can determine the correlation coefficient for two variables by using the Excel function =correl(range1,range2). For determining correlation coefficients between sets of variables, it's easier to use Data/Data Analysis/Correlations.

6) Now create the scatter diagram of revenue versus visits. (Given your previous work, what do you expect this plot to look like?) Determine the correlation coefficient of revenue and visits.

7) Summarize your results. In particular, elaborate on the implications of the relationship between revenue and number of visits to the website. Feel free to examine any other variable pairs you think might be important.

8) QA is interested in modeling data critical to their business. For example, if data for a particular variable appears to be reasonably approximated by a normal distribution, with a predictable mean and standard deviation, future values for that variable can be reasonably estimated. The purpose of the following exercise is to pursue this modeling process.

The Lbs. Sold worksheet contains the pounds of material sold per week from January 3, 2005, through the week of July 19, 2010.

> a) Determine the following summary values for this data: mean, median, standard deviation, minimum, and maximum.
> b) Create a histogram of the pounds of material sold data.

Excel Tip: Histograms. You can find a good tutorial on creating histograms in Excel at http://www.stat.ncsu.edu/people/reiland/courses/st101/1-4_excelhisto/sld001.htm.

Some suggestions:

Use the approach given in slide five as an estimate for the class width—you may want to experiment with smaller or bigger values. (Another rule of thumb for determining the number of bins is to start with the square root of the number of observations, and adjust from there.)

This tutorial uses Excel 2003. Use the following modifications for Excel 2007/2010:

On slide seven, choose Data/Data Analysis (not Tools/Data Analysis).

Replace the steps on slides 13–15 with the following (to reduce the gaps between the bars to zero): right-click on a bar, choose Format Data Series, set Gap Width to zero, and click Close.

The "fine tuning" suggestions in slides 18 on are optional. However, in Excel 2007/2010 you should outline each histogram bar in black: right-click on a bar, choose Format Data Series, Border Color, choose Solid Line, choose black for the color, and then click Close.

Another good (video) tutorial using Excel 2010 is at:
http://www.youtube.com/watch?v=RyxPp22x9PU.
Note that this approach uses two steps to better define the bin widths that appear on histogram.

---

c) Describe the histogram. Does it appear bell-shaped?

Generally, simply examining the histogram has limitations for judging how well a data set follows the normal distribution. (Bear in mind, for example, that while all normal distributions are unimodal and symmetric, not all unimodal and symmetric distributions are normal.) An additional approach is to see how well the data follows the Empirical Rule.

d) Determine how well this data follows the Empirical Rule by completing the following table.

Interval	Theoretical % of Data	Theoretical No. Obs.	Actual No. Obs.
mean ± 1 std. dev.	68%		
mean ± 2 std. dev.	95%		
mean ± 3 std. dev.	99%		

The theoretical number of observations refers to the number of observations in each interval if the data followed the Empirical Rule exactly. So, calculate the theoretical number of observations for each interval by multiplying the number of observations in the pounds sold data set by the theoretical percentage of data in that interval.

In order to determine the actual number of observations, you will need to count the number of observations in each interval. The best way to do that is to create a third column in the

---

Lbs. Sold worksheet that determines the z-score for each observation. Recall that the z-score is calculated as:

$$z = \frac{(observation - mean)}{standard\ deviation}$$

The z-score therefore gives the number of standard deviations an observation is away from the mean. (For example, if the mean and standard deviation of a data set are 25 and 5 respectively, then an observation of 35 is (35 – 25)/5 = 2 standard deviations away from the mean.) Once you've calculated the z-scores for the pounds sold data, sort the data by the z-scores. It then becomes straightforward to determine the actual number of observations within 1, 2, or 3 standard deviations of the mean.

> **Excel Tip: Counting.** If you've sorted your data by z-score, you can count which observations are in each category directly. Alternatively, you can use the Excel function =countif() to count all the entries greater than (>), less than (<), greater than or equal to (>=), or less than or equal to (<=) some value. For example, if you want to determine the number of observations in a range that are less than or equal to 1, you can use =countif(range,"<=1"). (Be sure to include the quotation marks.) You can count all the observations that are between 1 and -1 (inclusive) by using
>
> =countif(range,"<=1") - countif(range,"<-1").
>
> The Excel function =count() counts all entries in a range. You can therefore determine the proportion of observations less than or equal to 1 using =countif(range,"<=1")/count(range).

e) Refine your analysis by completing the following table for the pounds sold data.

Interval	Theoretical % of Data	Theoretical No. Obs.	Actual No. Obs.
mean + 1 std. dev.			
mean - 1 std. dev.			
1 std. dev. to 2 std. dev.			
-1 std. dev. to -2 std. dev.			
2 std. dev. to 3 std. dev.			
-2 std. dev. to -3 std. dev.			

f) How well does the data for pounds of material sold seem to follow the normal (bell-shaped) distribution? Support your response from your results in parts a through e. (I realize you don't have a standard here against which to assess "goodness of fit"—just use your best judgment.)

BY ROB WEITZ* AND DAVID ROSENTHAL†

Columbia
CaseWorks

g) Determine the skewness and kurtosis for the pounds sold data. Are these values consistent with your analysis of the pounds of material sold data?

> **Excel Tip: Skewness and Kurtosis.** In Excel, these values may be obtained via the =skew() and =kurt() functions. They are also included as part of the output when you determine the Descriptive Statistics for a range of data.

9) As part of the analysis, the number of *daily* visits to the QA website over the period May 25, 2008–August 29, 2009, was also collected. The material below is the output for the daily visits data using the same analysis you did in the previous problem for the pounds of material sold per week data. Your task in this problem is to use the two sets of output to write a paragraph or two comparing the distribution of the pounds sold data with that of the daily visit data. That is, is one more "normal" than the other? How do you know? (Note that the daily visit data is provided in the Quality Alloys Data spreadsheet; however, you may rely here solely on the output below.)

ANALYSIS OUTPUT FOR THE NUMBER OF DAILY VISITS TO THE QA WEBSITE OVER THE PERIOD MAY 25, 2008–AUGUST 29, 2009

Summary statistics using Excel's Data/Data Analysis/Descriptive Statistics command:

Visits	
Mean	150.2835498
Standard Error	4.567258109
Median	122
Mode	91
Standard	98.16949192
Sample Variance	9637.249143
Kurtosis	5.861989632
Skewness	2.166972212
Range	627
Minimum	37
Maximum	664
Sum	69431
Count	462

Histogram of Daily Visits
May 25, 2008 - August 29, 2009
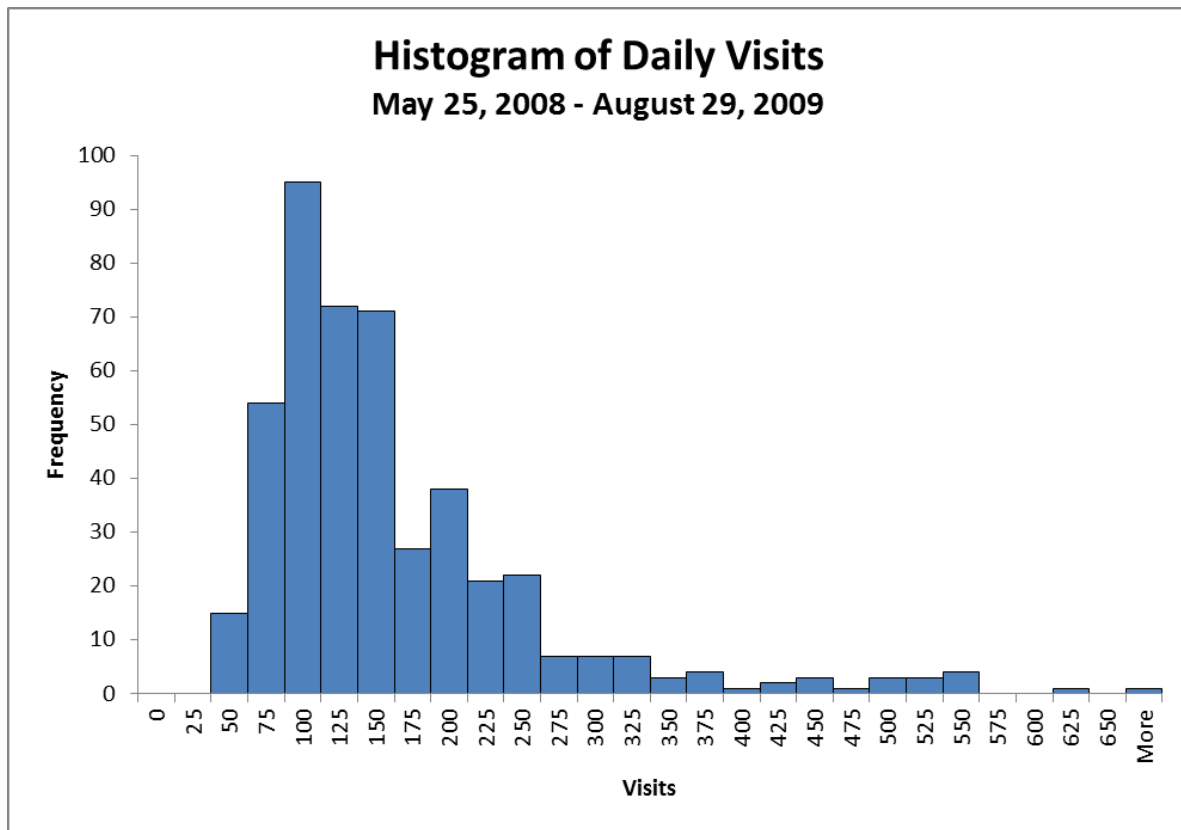
EMPIRICAL RULE AND MORE DETAILED ANALYSIS

Interval	Theoretical % of Data	Theoretical No. Obs.	Actual No. Obs.
mean ± 1 std. dev.	68%	314	392
mean ± 2 std. dev.	95%	439	439
mean ± 3 std. dev.	99%	457	448

Interval	Theoretical No. Obs.	Actual No. Obs.
mean + 1 std. dev.	157	108
mean - 1 std. dev.	157	284
1 std. dev. to 2 std. dev.	62	24
-1 std. dev. to -2 std. dev.	62	23
2 std. dev. to 3 std. dev.	9	9
-2 std. dev. to -3 std. dev.	9	0

The skew and kurtosis values are provided in the descriptive statistics calculations, above.

BY ROB WEITZ* AND DAVID ROSENTHAL†

Columbia
CaseWorks

SUMMARIZING DATA/SUMMARIZING DATA GRAPHICALLY

The data in the Demographics worksheet is included to (a) more fully provide you with a sense of the type of web data that can and is collected and (b) give you with a more complete picture of QA's customers and its website.

10) Represent each set of data graphically. In each case, write a sentence or two capturing the main conclusion(s) you draw.

Glossary

The following definitions are excerpted from the Google Analytics Glossary (http://www.google.com/support/analytics/bin/topic.py?hl=en&topic=11285).

BOUNCE RATE

Bounce rate is the percentage of single-page visits or visits in which the person left your site from the entrance (landing) page.

COOKIE

A small amount of text data given to a web browser by a web server. The data is stored on a user's hard drive and is returned to the specific web server each time the browser requests a page from that server.

Cookies are used to remember information from page to page and visit to visit, and can contain information such as user preferences or shopping cart contents, and can note whether a user has logged in so that they do not need to authenticate again as they navigate through the site.

NEW VISITOR

Google Analytics records a visitor as 'new' when any page on your site has been accessed for the first time by a web browser. This is accomplished by setting a first-party cookie on that browser. Thus, new visitors are not identified by the personal information they provide on your site, but are rather uniquely identified by the web browser they used.

NO REFERRAL

The "(no referral)" entry appears in various Referrals reports in the cases when the visitor to the site got there by typing the URL directly into the browser window or using a bookmark/favorite. In other words, the visitor did not click on a link to get to the site, so there was no referral, technically speaking.

PAGE

Any file or content delivered by a web server that would generally be considered a web document. This includes HTML pages (.html, .htm, .shtml), script-generated pages (.cgi, .asp, .cfm, etc.), and plain-text pages. It also includes sound files (.wav, .aiff, etc.), video files (.mov, etc.), and other non-document files.

Image files (.jpeg, .gif, .png), javascript (.js), and style sheets (.css) are generally not considered to be pages.

PAGEVIEW

A pageview is an instance of a page being loaded by a browser.

REFERRALS

A referral occurs when any hyperlink is clicked that takes a user to a new page of file in any website—the originating site is the referrer.

When a user arrives at your site, referral information is captured, which includes the referrer URL if available, any search terms that were used, time and date information, and more.

BY ROB WEITZ* AND DAVID ROSENTHAL†

Columbia CaseWorks

REFERRER

The URL of an HTML page that refers visitors to a site.

RETURNING SESSIONS

Returning Sessions represents the number of times unique visitors returned to your website during a specified time period.

RETURNING VISITOR

Google Analytics records a visitor as 'returning' when the _utma cookie for your domain exists on the browser accessing your site.

SESSION

A period of interaction between a visitor's browser and a particular website, ending when the browser is closed or shut down, or when the user has been inactive on that site for a specified period of time.

For the purpose of Google Analytics reports, a session is considered to have ended if the user has been inactive on the site for 30 minutes. You can update this setting with an addition to our tracking code.

TRACKING CODE

The Google Analytics tracking code is a small snippet of code that is inserted into the body of an HTML page. When the HTML page is loaded, the tracking code contacts the Google Analytics server and logs a pageview for that page, as well as captures information about the visit and non-identifying information about the visitor.

UNIQUE VISITOR SESSION

A Unique Visitor Session is a visitor interaction with a website for which the visitor can be tracked and declared with a high degree of confidence as being unique for the time period being analyzed.

UNIQUE VISITORS (OR ABSOLUTE UNIQUE VISITORS)

Unique Visitors represents the number of unduplicated (counted only once) visitors to your website over the course of a specified time period. A Unique Visitor is determined using cookies.

VISITOR

A Visitor is a construct designed to come as close as possible to defining the number of actual, distinct people who visited a website. There is of course no way to know if two people are sharing a computer from the website's perspective, but a good visitor-tracking system can come close to the actual number. The most accurate visitor-tracking systems generally employ cookies to maintain tallies of distinct visitors.

Endnotes

[1] Quality Alloys, Inc. is a pseudonym for the actual company represented here. Financial data (e.g., sales figures) have been disguised as have some details not relevant to the thrust of the case. The problem description and all other data are authentic.

[2] Laura Ramos, "Redefining B2B Marketing Measurement: Time for Business Marketers to Embrace Customer-Centric Metrics," Forrester Research, August 10, 2007.

[3] G. D. Bhatt and V. Grover, "Types of Information Technology Capabilities and Their Role in Competitive Advantage: An Empirical Study," *Journal of Management Information Systems* 22, no. 2 (2005): 253–277. R. Santhanam and E. Hartono, "Issues in Linking Information Technology Capability to Firm Performance," *MIS Quarterly* 27, no. 1 (2003): 125–153

[4] N. Carr, "IT Doesn't Matter," *Harvard Business Review* 81, no. 5 (2003): 41–49.

[5] For a brief description of the Google AdWords program, see https://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=6084.

BY ROB WEITZ* AND DAVID ROSENTHAL†

Columbia
CaseWorks