

AVINASH KAUSHIK



Web Analytics 2.0

[THE ART OF ONLINE ACCOUNTABILITY
& SCIENCE OF CUSTOMER CENTRICITY]



SERIOUS SKILLS.

INSIDE
Your Google AdWords™
Gift Card Worth \$25

Chapter 4. The Awesome World of Clickstream Analysis: Practical Solutions..... 1

Section 4.1. A Web Analytics Primer..... 2

Section 4.2. The Best Web Analytics Report..... 11

Section 4.3. Foundational Analytical Strategies..... 13

Section 4.4. Everyday Clickstream Analyses Made Actionable..... 20

Section 4.5. Reality Check: Perspectives on Key Web Analytics Challenges..... 52

The Awesome World of Clickstream Analysis: Practical Solutions

4

It's time to get down to real work. Yeah!

In this chapter, you'll learn about some of the most important web analytics reports, and I'll cover how to apply cool and effective reporting for SEO, site search, widget analytics, and more.

Then you will move to the next level of analysis ninjafication. You'll learn how to deal with some of the most complex challenges that bedevil our lives as people of data on the Web.

If you have skipped directly to this chapter, I do recommend at least skimming Chapter 3, because it's foundational to this one.

Chapter Contents

- A Web Analytics Primer
- The Best Web Analytics Report
- Foundational Analytical Strategies
- Everyday Clickstream Analyses Made Actionable
- Reality Check Perspectives on Key Web Analytics Challenges

When you crack open XiTi (or Omniture or your BFF analytics application), you'll face a bewildering array of reports. You'll start your lesson here by sharpening your skills in figuring out what's important and how to go about extracting value from your data.

A Web Analytics Primer

Web Analytics 2.0 is neither child's play nor impossible. Like everything in life, it is a complex journey. You can start simple and earn a white belt with easier tools, and then as you learn more, you can move on to more complex analysis and get a brown belt. Over time, as you get more experience and practice, you can work yourself to the all-powerful analysis ninja level!

With that in mind, we'll start with the understanding that you might be a *beginner ninja*, and with each recommendation, you take a path toward becoming more adept in your analytical strategy.

These recommendations are simple things that you can use on any website. My hope is that on a scale of ninjafication from a 0 to 100 (with 100 being the uber-master guru ninja), this section will move you from 0 to 35 in a matter of days.

I'll use an easy-to-understand structure for each recommendation:

- *What is it?* Understand what's in the report.
- *What is it telling you?* Know how to interpret the metrics and information.
- *What do you do next?* This is my little bonus gift for you; it's the next level of sophistication.
- *What is the bottom line?* Know what you should expect in the end.

Getting Primitive Indicators Out of the Way

Get ready to wade into some sweet numbers. Figure 4.1 shows you a snapshot of the key metrics you'll see when you log into a typical web analytics tool.

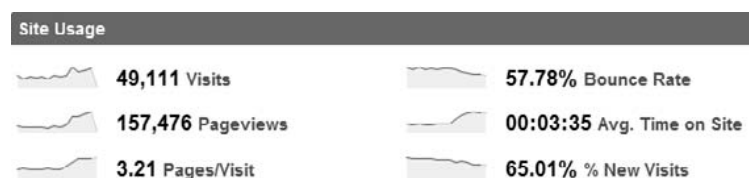


Figure 4.1 Base website key performance indicators

Because of their foundational nature, these six metrics—Visits, Bounce Rate, Page Views, Pages/Visit, Avg. Time on Site, and % New Visits—represent a great starting point for your web analytics journey.

Foundation metrics: what are they? Visits represent the number of sessions on your website, which is the number of times someone interacted with your site. Bounce Rate is the

number of those people who left instantly! Chapter 3 has more detail on each of these metrics.

The Page Views number is how many pages were requested in those visits. Oh, and Pages/Visit represents how many pages were requested in each visit.

Avg. Time on Site represents how long people stayed on your site.

% New Visits shows the number of sessions from people who visited your website for the first time.

What are they telling you? First, bask in the glory of how good you are (or be sad at how low your traffic is!). See those spark lines next to each number? They are trends for individual metrics and show a trend over time of that metric's performance.

Here is how I analyze the numbers you see in Figure 4.1, and you'll do something similar for your website: Visits look like they are all heading in the right direction for the business. It looks like an event a few weeks ago caused a peak, so I make a note to investigate that.

Next, I get clues that we are doing something very right because Pages/Visit and Avg. Time on Site are seeing a nice upward trend. You'll actually be surprised that frequently those two numbers do not correlate: people see lots of pages, can't find what they want, and spend little time. Or vice versa.

The Bounce Rate at 57 percent is definitely a concern, but recent changes to the site and traffic acquisition strategy have had a positive effect and reduced the Bounce Rate (see that spark line next to the number 57.78 percent?).

Growth of every business (for-profits and nonprofits) relies on growing the franchise by attracting new prospects. Although 65.01 percent of New Visits might look positive, notice that the spark line is trending down, which means although the overall number is a healthy 65 percent, we are getting more repeat visits (which would explain fewer bounces and more time on the site).

At the end of this quick analysis, I do my happy dance.

What do you do next? Dive into detailed trends. Beyond the quick spark lines snapshot, see how the metrics look over the last few months, and compare metrics between this month and last month.

No matter what web analytics tool you use, these comparisons will require clicking two or three buttons at most.

The bottom-line demystification #1 With a tiny amount of effort (30 minutes or so), you have learned your core metrics, and you know how you are doing on the surface. That wasn't hard, right?

Working through the primitive metrics gave you 5 out of the 35 points of progress toward greater ninjafication. Feel good? Yes!

Understanding Visitor Acquisition Strengths

It is critical to understand, early in your journey, where Visitors come from (you'll learn a bit more about this in the section that immediately follows). Figure 4.2 illustrates a typical Visitor acquisition report at a high level, which is all you need for now.



Figure 4.2 Major Visitor acquisition methods

A whole lot of excitement is packed in this humble little pie.

Visitor acquisition methods: what are they? Direct Traffic represents all those people who show up at your website after typing in the URL of your website or using a bookmark.



Note: It's important to code your campaigns with the right parameters to ensure that traffic does not show up in Direct Traffic.

Referring Sites represents other websites that link to you, including blogs, industry association sites, forums, competitors, your mom's site that proudly links to you, and so on.

Search Engines, well, that's you know who: Google, Yahoo!, Bing, Ask, and others. This bucket will include both your organic as well as your paid (PPC/SEM) traffic, so be aware of that.

Finally, there is Other. This bucket contains your display banner ads, email campaigns, social media campaigns, affiliates, and so on. Typically you are spending money with these places to acquire traffic (except for search).

What are they telling you? Look at the Direct Traffic first so you know how much traffic you're getting from people who already know your URL or have you bookmarked. Twenty-one percent is a healthy number for Direct Traffic. The best number will depend on the type of business you have, but a low Direct Traffic number could indicate issues with retention or truly connecting with customers.

Next, look at Referring Sites to identify sources that don't know you but are sending you traffic. You might visit the referring pages and see why. For some solid sites, you might want to establish a marketing relationship. Usually, referring sites link to you for free, and you want this number to be as high as possible, although it will never be 80 percent! Though, if this number is 2 percent, you may not be spreading your marketing message or value enough to get people to link to you. That is suboptimal.

The Search Engines bucket is very important. For many sites, search engines often account for at least a third and often half of the traffic. As more of the online population uses search engines to find information, this bucket is a key one to watch. In Figure 4.2, 60 percent is very healthy—though of course that depends on the business—but from my experience, I worry if I see, say, 10 percent. That’s just too low.

Typically, Other will be a small part of your pie, which is a good thing because Other contains your acquisition strategies, in other words, things you are spending money on to get traffic. If, for example, your Other bucket is 45 percent, that is a sign you might not be working hard enough on getting *free traffic* (Organic Search, Referring Sites, Direct Traffic).

What do you do next? Look at longer-term trends for each bucket. Are you getting better or worse over time? Consider segmenting the data and diving deeper in higher-priority areas. For example, Figure 4.3 shows a simple and quick approach to segmenting Search Engine traffic.



Figure 4.3 Segmenting organic and paid search traffic

My initial interpretation of the segmented data in Figure 4.3 would be that there is perhaps *overreliance* on paid search. So, you dive deeper to understand why. See how it works?

I also recommend drilling down to specific websites that send you traffic and, of course, drilling down to keywords and key phrases that are sending you traffic. Both of those help you understand that critical customer *intent*.

Oh, and in both cases, look for surprises.

The bottom-line demystification #2 You probably spent 30 minutes looking at the previous figure and drilling down to the next level of report. You now know where your Visitors come from, and you have a first-blush understanding of whether your company is spending its marketing dollars effectively.

You are now 10 points into your 35 points of progress of ninjafication! Hurray!

Fixing Stuff and Saving Money

You’ve spent an impatient hour working on some reports; now it’s time to fix stuff and save some money! First you must identify web pages that are a *key point of failure*. In

Chapter 3 I covered exactly how to do that: by examining your site's Top Entry Pages and the Bounce Rate for each.

Another immensely actionable way to find places to fix stuff (and save money) is to look at the top 25 keywords driving traffic to your website and their Bounce Rates (or as Figure 4.4 from Clicktracks calls it: Short Visits).

		Google	Windows Live	Yahoo!	MSN	Ask	AOL
Total	84.5%	83.2%	97.5%	88.3%	93.9%	91.2%	89.6%
avinash kaushik	49.2%	49.1%	64.7%	40.7%	100.0%	100.0%	66.7%
avinash	59.0%	48.2%	98.4%	34.1%	100.0%	-	100.0%
survey questions	94.4%	94.1%	100.0%	97.7%	94.1%	77.8%	100.0%
occam's razor	59.2%	58.8%	52.2%	73.3%	81.3%	-	100.0%
analytics	99.7%	89.2%	100.0%	-	-	100.0%	-
working at google	86.2%	86.2%	-	88.9%	-	-	66.7%
4q	59.8%	60.1%	-	43.8%	-	-	50.0%
kaushik	66.0%	59.3%	96.7%	66.7%	100.0%	-	0.0%
working for google	87.9%	87.6%	100.0%	93.3%	100.0%	100.0%	100.0%
web analytics tools	72.3%	72.5%	0.0%	-	-	-	-
occams razor	55.7%	55.7%	44.4%	50.0%	60.0%	-	100.0%
customer survey questions	93.5%	93.7%	100.0%	100.0%	92.8%	100.0%	0.0%
kpi	89.0%	89.0%	100.0%	100.0%	-	-	-
web analytics blog	53.5%	53.7%	-	0.0%	-	-	-
google trends	78.0%	78.1%	-	-	-	-	71.4%
google	96.6%	89.6%	100.0%	-	-	100.0%	100.0%
key performance indicators	89.8%	89.3%	-	100.0%	-	-	-
occam's razor blog	49.6%	48.9%	100.0%	60.0%	100.0%	-	-
tracking cookies	91.5%	92.7%	77.8%	62.5%	100.0%	-	100.0%
data driven decision making	91.7%	91.3%	-	-	-	100.0%	100.0%

Figure 4.4 Percent Short Visits (bounces)—Clicktracks

Top entry pages, keywords, and bounce: what are they? Here's a metaphor: you have many doors into your department store. With Top Entry Pages and Bounce Rate, you identify doors that are not letting people come into your store. Talk about a crime!

The Top Entry Pages report shows you the top home pages that let your Visitors, and precious little you, down. The keyword report helps you identify keywords where something is amiss. It's even better if you look at just your paid search keywords.

What are they telling you? Simple: Stinkiness. Your site's.

Remember, Bounce Rate measures stinkiness from a customer's perspective: "I came, I puked, I left." Pages with a high Bounce Rate are not delivering on the promise that drives customers to your site. The ones in the top ten entry pages report need your attention. You fix 'em, and you increase the likelihood that people will go deeper into your site and maybe convert.

The keywords report is even more interesting. Here you have *intent*. The customer is telling you why they might be visiting, and keywords with high Bounce Rates are where you are not meeting that intent. You may be ranked for the wrong keywords, or the pages these folks land on may not have the right calls to action. Fix it.

What do you do next? I'll discuss this in some detail in Chapter 7, but your real friend here is Experimentation and Testing. Start with simple A/B tests. Start with a free tool like Google Website Optimizer; it can do 95 percent of what any paid tool can do. If you already have a paid tool like Optimost, Offermatica, or SiteSpect, then go for it with one of those.

Pick pages you want to fix, create a couple versions of the pages, and put them into a test. Change copy, content, images, and calls to action—everything is fair game.

With testing, you improve the pages based on customer feedback.

The bottom-line demystification #3 It takes you about two minutes to get to each report and another minute to look at the numbers and click a few buttons. At the end of the half hour, you have created a specific list of focus areas. You know exactly where to start fixing pages on your site and potentially improving your paid campaigns.

You've spent an hour and a half, and you are up to 20 of your 35 points in the analysis ninjafication process. And now you know what you are doing! Priceless!

Click Density Analysis

I love the site overlay report. There, I said it. I love it.

You have improved the Top Entry Pages and key traffic-driving campaigns (key phrases). Now figure out why pages that you want to win on your site are not winning and why pages with key calls to action are not delivering. Look for the obvious things you are doing wrong. Figure 4.5 is the Google Analytics site overlay report for www.fotonatura.org.

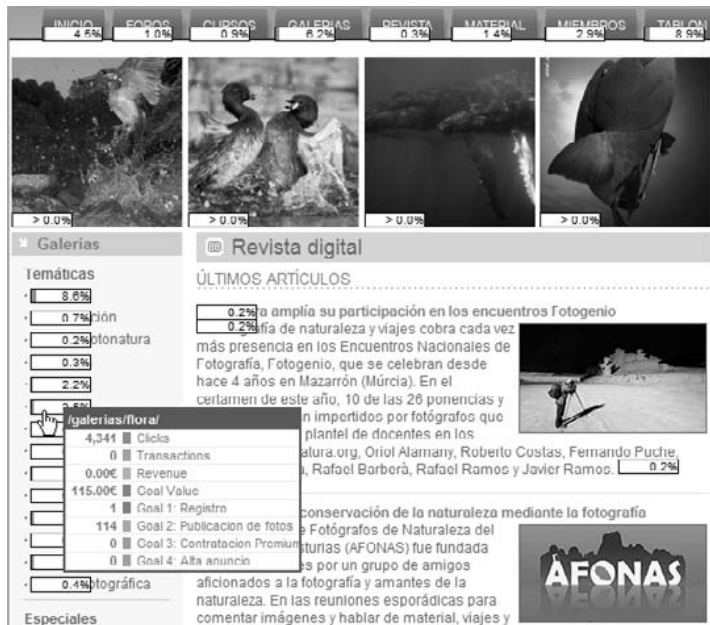


Figure 4.5 Site overlay report—Google Analytics

There are two reasons for my fondness for this report:

- For many people, numbers, metrics, and spreadsheets are still overwhelming. A site overlay report demystifies all that. You see the data visually represented.
- Even seasoned analysts are not as good at analysis as they should be because they rarely use the website they are analyzing. Using the site overlay report is a great way to walk in the shoes of the customers.

Site overlay: What is it? Click density analysis.

The site overlay report shows the number of clicks on each link on a page (as in Figure 4.5). It also shows you other helpful information such as Revenue from clicks on a link, Goal Conversions, and so on.

Each tool will show click density slightly differently. My favorite may be ClickTracks because it also shows context and key metrics about the page, right in the site overlay report, as you can see in Figure 4.6.

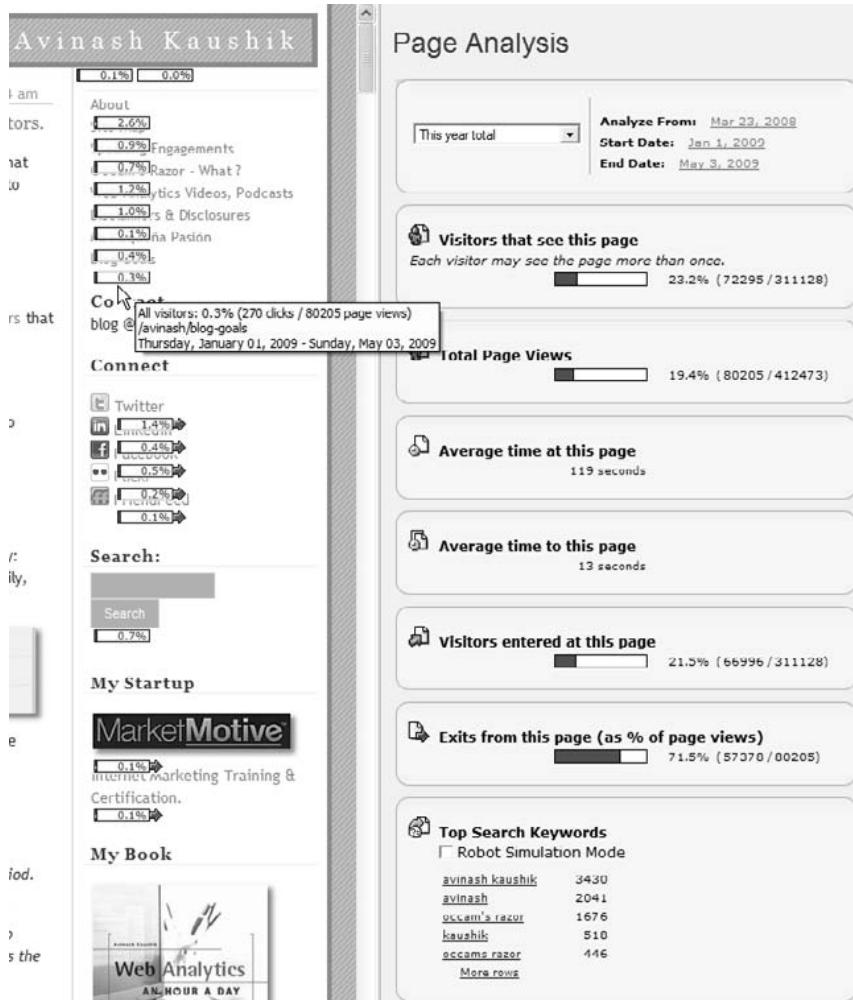


Figure 4.6 Navigation report with page-level metrics—ClickTracks

The site overlay report shows % Page Views, Time on Page, Time to the Page, % Exits, and Keywords that brought people to the page. In a nutshell, it's everything you would ever want to know to judge the performance of a page.

What is it telling you? When you look at the site overlay report, you are looking for clusters of heavy clicks. Look for the top two or three most clicked links, and try to reconcile that against links that you want Visitors to click. See what people are clicking “below the fold,” and look for any surprises there.

Also, look at links that ultimately drive high conversions (you can have conversions on an ecommerce website or, as in Figure 4.6, a non-ecommerce website). Look for things that connect with people. For example, do more people convert on the site when they click Product Comparison on the home page or when they go directly to a product page? Try to follow a few heavy clicks and see what people do next. Walk in their shoes, and experience your own website.

Check out referrers to each page; that could explain Bounce and Exit Rates.

Finally, look at Average Time to This Page. If it takes too long for people to find your key pages, then you have a problem with your core site navigational elements or merchandizing.

What do you do next? Identify improvements to your pages.

Consider merchandizing and cross-sell and up-sell opportunities now that you know what people like. For example, no one is clicking your blinking promotion in the middle of the page because it looks like an ad!

If your tool allows, segment the clicks. Where do people who convert click vs. everyone else or vs. everyone from a search engine or an email campaign?

Also, by segmenting the data, you not only understand things in aggregate, all Visitors, but now you can start to understand different types of people on your site. And now you can treat them differently!

The bottom-line demystification #4 Looking at the site overlay report is a very visual, easy-to-understand way to learn exactly how people browse your website, which methods are working on your site, and which are not. No tables, no numbers, no graphs—even your HiPPO will get this!

You can take up as much time as you have available on the site overlay report. Initially, you'll probably spend 30 to 60 minutes exploring your top pages. At the end of that, you'll be up to 25 out of 35 points of ninjafication!

Measuring Visits to Purchase

There is a unique phenomenon that many of us dismiss: people usually don't buy on the first visit. But most of us focus on getting people to buy right away, especially those of us who measure Conversion Rate as Outcomes or Visits.

The antidote to that suboptimal mental model is the Visits to Purchase report, as shown in Figure 4.7.

Visits to Purchase	Transactions	Percentage of all purchases
1 visits	237	42.70%
2 visits	106	19.10%
3 visits	72	12.97%
4 visits	49	8.83%
5 visits	26	4.68%

Figure 4.7 Visits to Purchase report

Visits to Purchase: what is it? Your web analytics tool starts anonymously tracking a visitor from their first visit. When they purchase something from your website, that event is noted. The Visits to Purchase report shows the distribution of the number of visits it took for someone to purchase from your website. A “purchase” can also be a submission or a lead or another such outcome. For example, if President Obama is collecting email and contact information of people interested in supporting his reelection in 2011, then his staff can track Visits to Lead submissions. Same method.

What is it telling you? You are looking for how evenly the numbers are distributed. In Figure 4.7, only 42 percent of the people purchased on the same visit. Then there is a nice drop-off that continues to xx visits.

You use this information to help you understand the intensity of the *pan-session* purchase behavior of your customers. You can learn what it takes to convince people to purchase the product you are selling. If you sell iPods and Ferraris on your website, it is incredibly valuable to understand that most people purchase Ferraris on the first visit, while they take 15 visits to buy an iPod. (What can you say? Midlife crises cause such odd behavior.)

What do you do next? The *sister* report of Visits to Purchase is Days to Purchase. The intersection of these two reports helps you understand how many days pass between visits. For example, most people may make a purchase after three visits to your site. But those three visits could be more than fifteen days. Or it could be that those three visits were on the same day.

This analysis helps you understand customer behavior in a very actionable way. Now you can go back and optimize how you sell each item and how you advertise and market it, and you can even optimize your inventory system!

The bottom-line demystification #5 Reviewing the Visits to Purchase and Days to Purchase reports and understanding the data are not very time-consuming, and the reports are standard in most tools. The process may take, say, 20 minutes. It will take a bit more time, maybe the rest of that hour, to correlate the data and cross tab that with your selling strategy and find actionable insights.

Regardless, at the end of the hour, you will have earned all 35 points! You are now officially well on your way to uber-master guru ninja! High-five!

Consider this: in approximately three hours of work, you went from knowing a little bit about web analytics to a starter ninja, and along the way you were already taking actions based on the lessons you learned. Congratulations.

The Best Web Analytics Report

At pretty much every conference I present, I get asked this question: “If you could pick only one web analytics report to take with you on a deserted island, which report would it be?”

I can do a cop-out: well, it really depends on your business. What are the three strategies you are executing? Are you B2B or B2C? You catch my drift.

Those are tough questions, but you can’t get paralyzed just because you don’t know enough (an analysis ninja understands one can’t ever have complete information).

There is one report that will work for any type of website, and it qualifies as my nominee for the best web analytics report: *Outcomes by All Traffic Sources*. Coincidentally, it also fits perfectly with the Web Analytics 2.0 core mental model and hence is immensely actionable. Figure 4.8 shows how an Outcomes by All Traffic Sources report looks, in this case in Google Analytics.

Site Usage		Goal Conversion				Views: [Table Icon] [List Icon] [Full Screen Icon]		
Visits	Goal1 Conversion Rate	Goal2 Conversion Rate	Goal3 Conversion Rate	Goal4 Conversion Rate	Goal Conversion Rate	Per Visit Goal Value		
29,111 % of Site Total: 100.00%	1.20% Site Avg: 1.20% (0.00%)	2.17% Site Avg: 2.17% (0.00%)	0.66% Site Avg: 0.66% (0.00%)	0.96% Site Avg: 0.96% (0.00%)	4.99% Site Avg: 4.99% (0.00%)	\$0.50 Site Avg: \$0.50 (0.00%)		
Source	Visits ↓	Goal1 Conversion Rate	Goal2 Conversion Rate	Goal3 Conversion Rate	Goal4 Conversion Rate	Goal Conversion Rate	Per Visit Goal Value	
1. google	10,809	1.31%	2.07%	0.80%	0.70%	4.89%	\$0.50	
2. (direct)	7,597	1.26%	2.36%	0.46%	1.09%	5.17%	\$0.50	
3. google.com	1,575	1.08%	1.71%	0.95%	0.70%	3.94%	\$0.46	
4. images.google.com	583	0.00%	0.00%	0.17%	0.00%	0.17%	\$0.03	
5. twitter.com	560	0.18%	3.21%	1.25%	0.89%	5.54%	\$0.56	
6. stumbleupon.com	469	0.00%	0.00%	0.00%	0.00%	0.00%	\$0.00	
7. analytics.blogspot.com	453	1.32%	2.65%	0.66%	0.60%	5.30%	\$0.50	
8. yahoo	389	0.26%	1.03%	0.77%	0.51%	2.57%	\$0.31	
9. wilsonweb.com	223	0.90%	6.28%	0.45%	0.45%	8.07%	\$0.56	
10. googleblog.blogspot.com	195	1.54%	1.03%	0.51%	1.03%	4.10%	\$0.46	

Find Source: containing [] Go [Go] Go to: 1 Show rows: 10 1 - 10 of 1,041

Figure 4.8 Outcomes/Conversions by All Traffic Sources

In a veritable ocean full of metrics in our web analytics tools, this report represents two things you should care about more than anything else: sources of traffic and Outcomes.

Sources of Traffic

I have come to believe that if you know the source of your traffic, then you can strongly infer what kinds of people are coming to your website and even a little bit of why they are coming (intent).

As an example, I am a huge fan of Direct Traffic, and I will talk about that more later in the chapter. If you tag your campaigns correctly, then Direct Traffic represents free traffic because it comes from people who arrive via using bookmarks, typing in your URL, or other such activities. Direct Traffic is also traffic that is familiar with you, so it typically represents returning Visitors and most likely your existing customers. In Figure 4.8 I was happy that Direct Traffic was so big—and notice that it converts higher, which is very typical.

But if during the last 30 days I had spent a boatload of money trying to attract new Visitors (*prospects*) to my website, a big number from Direct Traffic might not be such a good thing.

Notice the box drawn around twitter.com in Figure 4.8. In the prior 30 days, it was not even in the top 20, and now it's at 5. Immediately you can see how social media efforts might be paying off. You can also infer that these are much more tech-savvy people; in fact, they're the bleeding-edge kind. The source helps you better understand the persona of the audience.

Row 6, stumbleupon.com, represents new Visitors who might typically be interested in recent stories, or if this were an ecommerce website, it would represent recent promotions, product launches, and so on. Stumbleupon.com, digg.com, and others also represent a sense of validation that your content is good and it is being spread by others whom you don't know.

Row 9, wilsonweb.com, represents traffic that I would otherwise never reach, namely, very traditional direct response and traditional marketers. Dr. Ralph Wilson has written about the Web since 1995 and reaches a unique audience—and I am grateful to him for sending me free, valuable traffic.

Finally, the importance of Google to this website is very clear. The numbers validate that the work put into search engine optimization (SEO) is paying rich dividends. A specific example of this is row 4, which is traffic from images.google.com; that is a validation of the time and effort spent tagging each image on the website with relevant descriptions.

These examples should show you how much you can learn by understanding the sources that send you traffic. You can see what's working and what's not in terms of your core acquisition strategy and whether you are attracting the right audience to your site.

Now should these sites be on your *permanent BFFs* list? That depends on if they send you quality traffic, and here we go....

Outcomes

In Chapter 1, I stressed the importance of Outcomes; that's the reason they are a key part of the best Web Analytics 2.0 report. If you want to change the culture of any company, you start by focusing on Outcomes (not measuring Visits, Time on Site, or Top Exit Pages—yikes!).

Figure 4.8 shows Conversion Rate for a non-ecommerce website. At a glance, I can see the overall Conversion Rate (the one with the box around it), and I can also go down the list of the websites and very quickly identify which ones are sending me quality traffic, in other words, traffic that takes action that adds to my bottom line.

I can also quickly identify my BFFs (twitter.com in this case). I can quickly see sources that send me lots of traffic yet are not my real BFFs (stumbleupon.com in this case, or even images.google.com). Notice how in one view, Visits, you come to different conclusions than looking at another, perhaps more important, view.

You can dive into many more nuances in this report. Notice there are multiple goals for you to consider. Each part of your business might have different Outcomes, or *goals*, which are shown in the report as Goal2, Goal3, and Goal4. The report will tell each business unit or leader how they are performing against their unique goal.

The essence of the best web analytics report is that it highlights two questions to focus on first: *who?* and *how much?*

If you start with the Outcomes by All Traffic Sources report, you'll find that your senior executives suddenly care about your web analytics reports. They will ask you good questions, and they will seek you out rather than you knocking on doors that never open.

Ain't that sweet?

Foundational Analytical Strategies

The rest of this chapter will help you evolve your thinking about Clickstream analysis (frequently called *web analytics*). You'll learn a few to-die-for analytical strategies such as segmenting and focusing on Customer Behavior metrics. Then we'll leap into specific types of analyses you'll do in your daily life.

Through each lesson, my hope is that you'll focus not just on the metrics and reports being analyzed but rather on the reasons for choosing the metrics and the thought process around creating insightful analyses. That'll ensure that other types of analyses you do, not covered in this book, are actionable.

Segment or Go Home

Absolutely nothing is more important in analytics than segmentation. Why? Because monoliths don't come to our websites. I know that seems like such a shock. Yet most of our reporting and analysis happens at an aggregate level. That's like saying Figure 4.9 represents the traffic that comes to our site.

The reality is that your traffic actually looks like Figure 4.10.

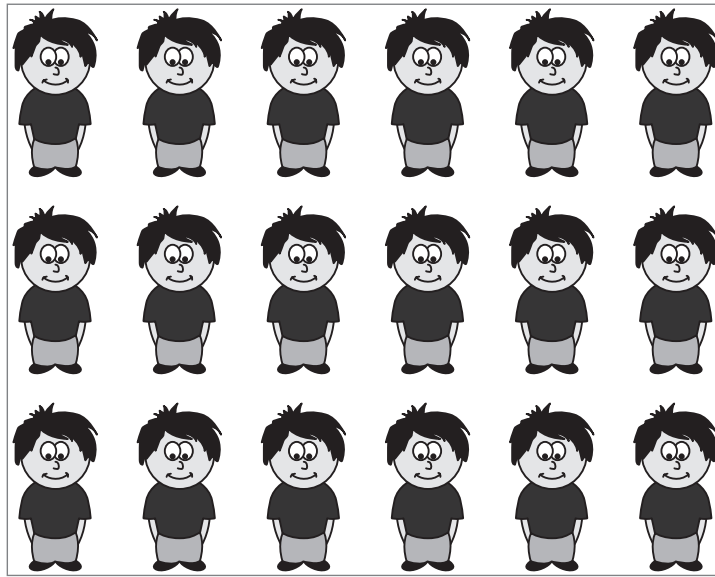


Figure 4.9 Representation of analysis traffic in aggregate

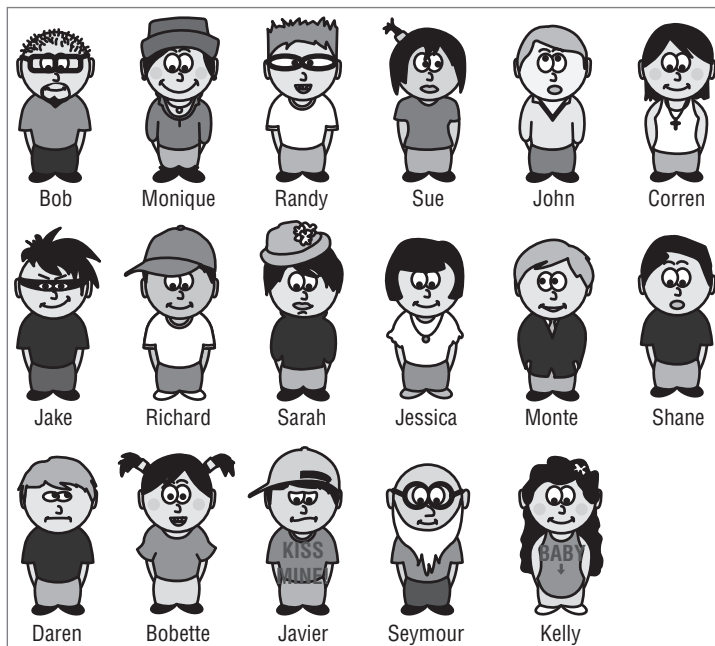


Figure 4.10 Actual reflection of your website traffic

Different types of Visitors come to your website. And they all come with different intentions, with different problems or solutions, and with different personas.

This unique phenomenon mandates that you have a very effective and persistent segmentation strategy as part of your web analytics process.

Benefits of Segmenting

The power of segmenting a metric is that you peek behind the curtain and find out more about the metric. These are the benefits that you will gain:

- It is impossible to create a segment of your data without putting in the effort to understand what is important to your business, what goals you want to accomplish. This means you'll have to spend time understanding the business, a good thing.
- By segmenting your data, you can quickly hone in on areas of deeper depth, which will reveal key insights that drive meaningful action.
- Our senior executives and decision makers don't understand all the complexity and magic of a web experience. Showing them segmented trends is an extremely effective communication tool. And the best part is you barely have to talk; the picture will tell the story!

Segmenting your data doesn't mean focusing on the *glob* but rather focusing on the *specific*. That focus helps make ideas actionable. To internalize the power of segmentation, let's look at an illustrative example.

Figure 4.11 shows revenue generated from a website over a period of 12 months. I am positive you have a graph like this running loose in your company.

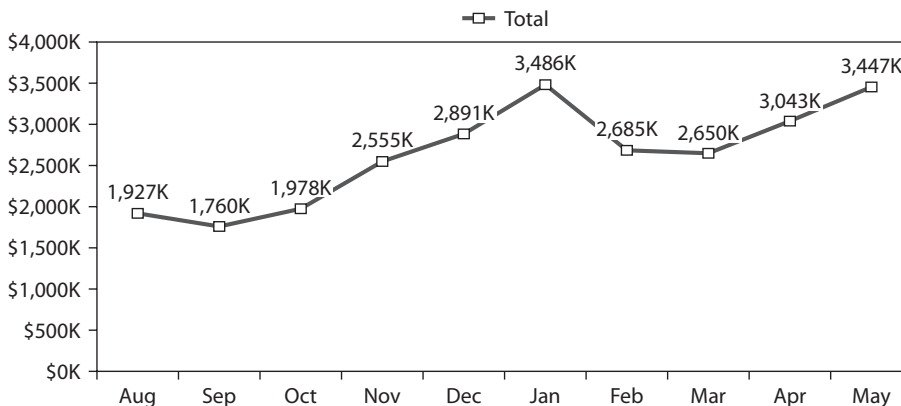


Figure 4.11 Monthly revenue trend over 12 months

Cute. You can see things going up and down. Your executives already know that December is always a seasonal high for the company, as is May. You have just shown them what they already know. The actual numbers are mildly interesting. But what decision can they make from this? Zilch. Nada. Zip.

Now try Figure 4.12, the same monthly trend, but this time illustrating important segments of the overall revenue.

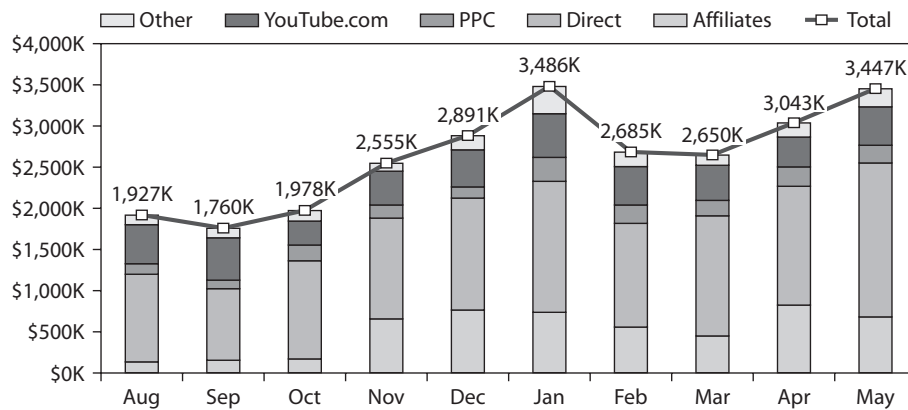


Figure 4.12 Segmented monthly revenue trend—sources of revenue

Sweetness! Your executives can now see a delightful monthly trend of the main contributors of revenue to the business. They can quickly conclude that for all the money going down the pipes, direct, *free traffic* dominates revenue contribution! The paid search, or PPC, program needs serious attention. Despite spending *muchos dineros*, it still does not deliver a proportional chunk of revenue. Oh, and look how a *rogue marketer* did a tiny experiment with YouTube. Now she deserves a bonus because her experiment has consistently produced revenue.

See what I mean? You can report the data, or you can effectively segment it and communicate more value to your decision makers.

Creating and Applying Segments

Although identifying the optimal behaviors, sources, metrics, and Outcomes to segment remains a skill you will develop over time (and ninjas *do* possess this skill), creating segments and applying them to your reports is pretty much a piece of cake.

Say I work at Intel and I am responsible for improving Intel's marketing in Eastern Europe. I could just segment www.intel.com's data using geographic segmentation. Or I could get way more specific and earn my salary by understanding customer behavior using segmentation.

Figure 4.13 illustrates my strategy. I open the folders on the left (under Group Selection) and drag and drop my metrics or dimensions onto the palette:

1. First I drag over Organic Search. I really want to know how I am doing with free search traffic.

2. Next, I am interested only in traffic that looks at content about microprocessors (sure, Intel does other things, but my bonus is tied to microprocessors!).
3. I am less interested in understanding what happens to traffic that bounces; I really want to know keywords that bring traffic that visits my site multiple times (Number of Visits greater than 3).
4. Finally, I'll restrict my analysis to just traffic from Eastern Europe.

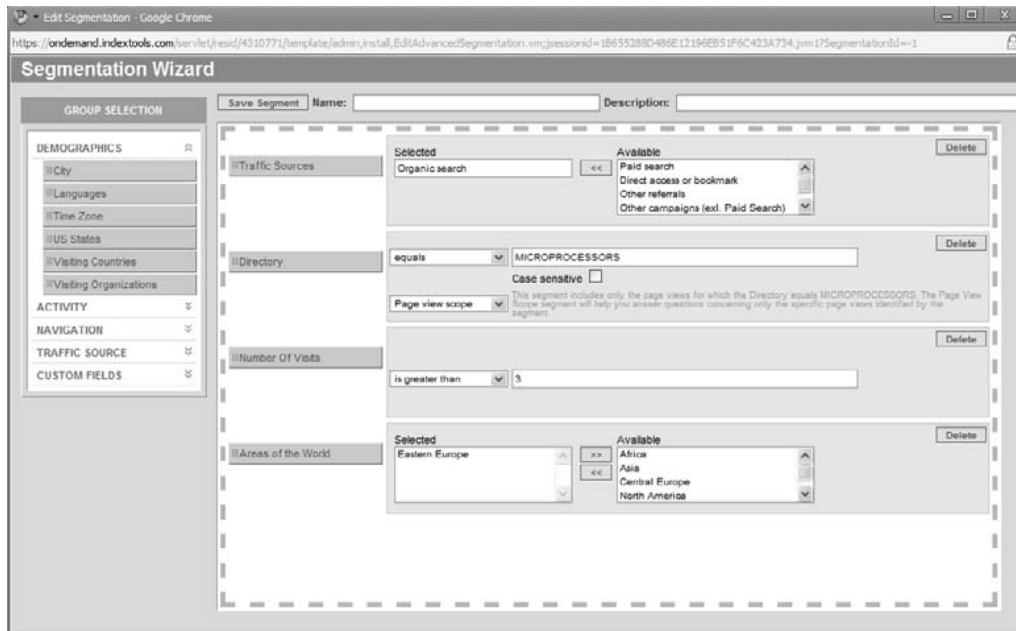


Figure 4.13 Deep segmentation of website data—Yahoo! Web Analytics

In the past, you had to know complicated SQL, and you had to know basic database design to segment this way. Now, you drag, you drop, and bam!

I can now go apply this segment to my content report, which helps me understand what kinds of microprocessors are most interesting to this precious traffic segment. I can apply it to the search reports and understand the search engine preferences of folks in Eastern Europe, as well as what valuable keywords send this kind of quality traffic. I can dive deeper to see whether they downloaded product specifications or submitted leads.

Here's a final example to illustrate how you can answer everyday questions that might seem to have complicated answers. My question was, how effectively does my home page engage Visitors so that they read lots of my content? Put another way, how good was my first impression?

That seems like a hard question, no? Figure 4.14 illustrates how I can get that answer very quickly. From the Dimensions and Metrics lists, I drag over the metrics I need: I take Page Depth first and then Visits where Visitors have seen more than three

pages (a high *degree* of engagement—see Chapter 3). Next I choose Landing Page, because I want only those people who start browsing at my home page and not people who entered deep in my site and then surfed over to the home page.

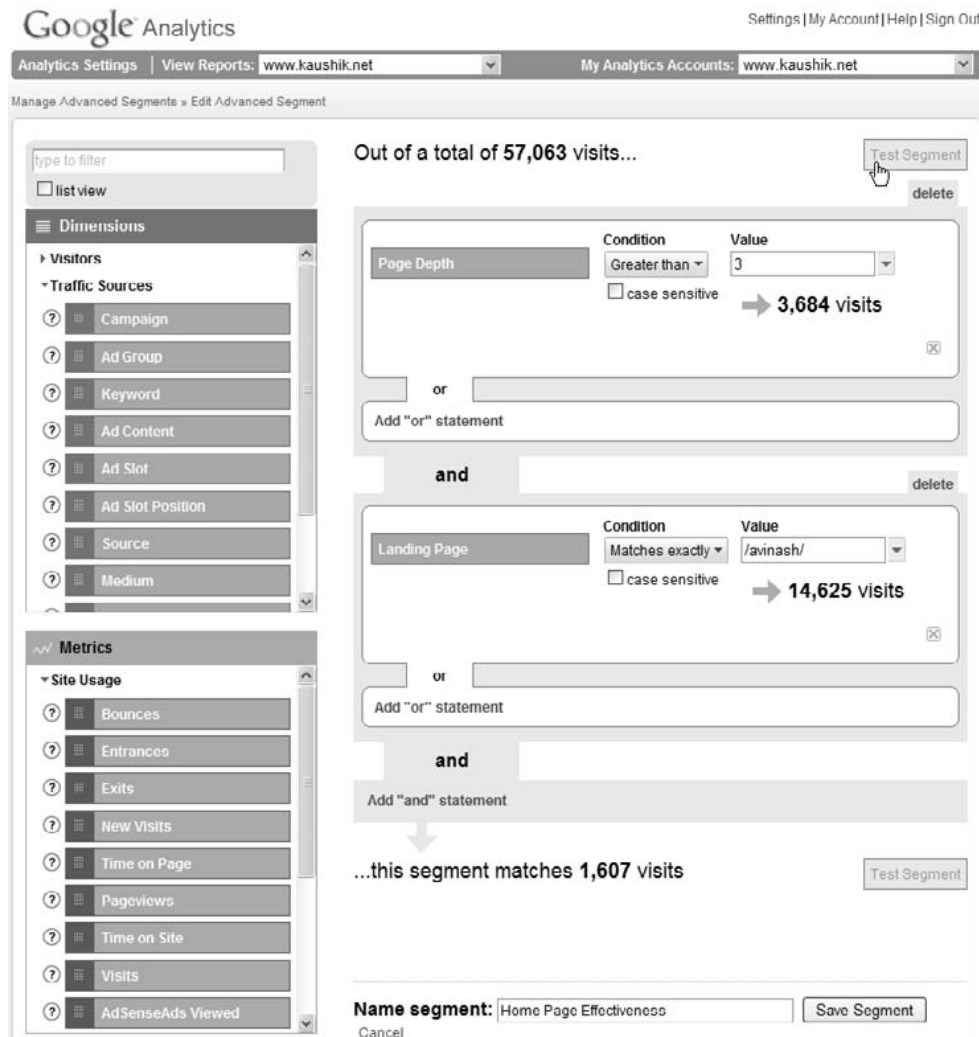


Figure 4.14 Home page effectiveness in driving a high degree of engagement—Google Analytics

In this time period, there were 57,063 Visits to my site. Of those, 14,625 entered the site on the home page (see what I mean when I say home pages are dead?). In total, 3,684 Visits had people reading more than three pages.

And my answer? 1,607. Only 1,607 out of a possible 14,625 thought my home page made a good impression! Me sad now. But do you see how simple it was for me to get the answer, once I knew my question?

That example illustrates a very cool feature in Google Analytics. In most web analytics tools, you must first tell your vendor what data you might want to segment,

and then they suggest implementing JavaScript tag updates. You wait a few days to get data, and then you must create the segment, submit it, wait for processing, and finally get the data you want.

In Google Analytics (and in Yahoo! Web Analytics), you don't have to tell Google up front what you want to segment, you don't have to change tags, and you don't even have to submit the segment and wait for it to be processed. You simply drag and drop, and when you are done, you click Test Segment (the top-right button in Figure 4.14), and you have your answer right there in your Advanced Segments creation window. Pretty yummy.

Now, you go do the same process with Google or Yahoo! Web Analytics and see how good your home page is.

To sum it up: you no segment, you no find insights. You no segment, you no have job for long time.

Segmentation is a strategy that has been applied thus far in this book, without you probably realizing it, and will be applied liberally for the rest of this book. My hope is that you'll use it just as liberally as you tackle the problem of answering tough business questions.

Focus on Customer Behavior, Not Aggregates

Aggregates again!? I hear you yell. Yes. But in this section, *aggregates* means something slightly different. I want to shed light on a different way of analyzing data. We will not focus on aggregate metrics such as Unique Visitors and Average Time on Site and such; rather, we will identify truly actionable insights and focus on metrics that report the behavior of customers on your website.

Measuring the number of Visits to your site during a month is interesting. Maybe a rise or fall in that number, or meeting a preset goal, will be of some value. What's more valuable is focusing on a *behavioral* metric like Bounce Rate because it measures, as you already know, this customer behavior: "I came, I puked, I left."

Analyzing your site Bounce Rates at various levels helps you understand what's working, what's broken, and what you need to do more or less (check out Chapter 3 for more specifics).

Let's look at an example. I am responsible for running the website `news.bbc.co.uk`. Success will typically be measured using an *aggregate* metric such as Page Views or Pages/Visit, as shown in Figure 4.15.



Figure 4.15 Three aggregate metrics: Visits, Page Views, Pages/Visit

There aren't too many insights there, sadly. The customer behavior that is of value to the BBC is people seeing more than four pages on the website. Figure 4.16 shows the distribution of Visits by the number of pages read during each visit.

Depth of Visit

Most people visited: 1 pages

Depth of Visit	Visits	Percentage of all visitors
1 pages	67,671.00	52.95%
2 pages	17,545.00	13.73%
3 pages	12,166.00	9.51%
4 pages	6,000.00	5.37%
5 pages	5,053.00	3.95%
6 pages	3,413.00	2.67%
7 pages	2,657.00	2.08%
8 pages	1,994.00	1.56%
9 pages	1,570.00	1.23%
10 pages	1,261.00	0.99%

Figure 4.16 Distribution of pages read during each Visit

Ahhh...now this is fantastic—and actionable. You can see that 76 percent of the website Visitors did not exhibit the ideal behavior. Now you can easily segment the behavior that is accretive to your business, four pages or more, and analyze customers (the 24 percent) who exhibit that behavior. You can dive in and understand what their preferences are. Do they like sports? Podcasts? International stories? You catch my drift.

This will help you understand what your valuable customers consume, which in turn helps you focus better. (Maybe dump entertainment? Or make it a lot more interesting?) Of course, you can create the inverse segment—that is, fewer than four pages—and see what that group is reading, where they come from, and what repulsed them.

Behavioral metrics are all around you. Loyalty, Recency, and Returning Visitors are just a few that come to mind. I'll talk more about these metrics in Chapter 5. Focus on behavior, and glory awaits you.

Everyday Clickstream Analyses Made Actionable

We are going to focus on five very common types of Clickstream analyses that you'll do when you have a bit more gray hair or, to be more politically correct, when you're more experienced: internal site search, search engine optimization, paid search, direct traffic, and email campaigns.

Internal Site Search Analysis

Almost all web analytics clicks data is missing one key ingredient: customer intent.

The keywords that people type into search engines such as Google, Bing, and Yahoo! contain a modicum of intent. The real gold is the search engine that you surely have on your website. What? No, you don't? You are the last website on the planet not to have an internal site search engine? Shame! Shame!

If you directly understand the intent of Visitors on your site, you can better understand the causes for success or failure on your site.

Here's an example. You can look at the top 10 most viewed pages on your site and understand what people who came to your site wanted. Or do you? How would you know which pages your Visitors *wanted* to see? If Visitors can't find those pages, then your web analytics tool won't record that action.

One way to overcome the challenge of intent is to look at your internal site search data and see what customers typed into your site search engine. You should perform three clusters of actionable analysis with your internal site search data: site search usage, site search quality, and, wait for it...segmenting.

Site Search Usage

When you look at how your internal site search is used, you need to answer the basic questions first: how much is the search function used, and what keywords are used most?

Figure 4.17 shows how a report that illustrates the trend for site search usage might look.

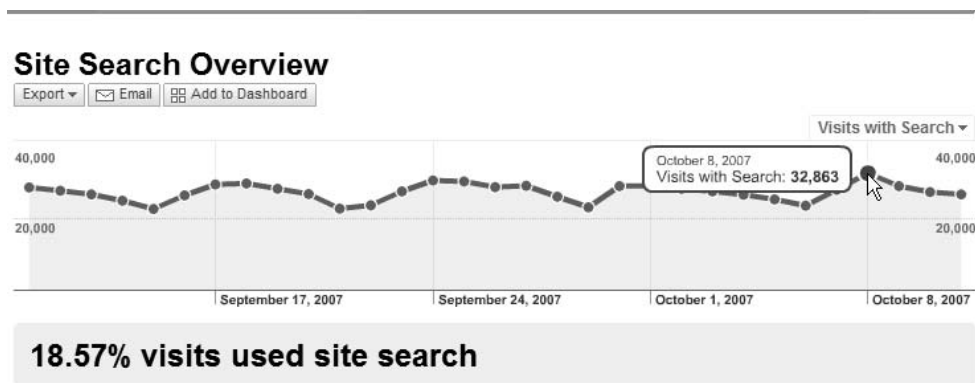


Figure 4.17 Trend for internal site search usage

At a glance, you can see how important site search is (18.57 percent) and how the trend tracks over time. Visitors will use no other navigational aid even a small percentage of the time; typically, site search will be the site navigational element they use most often.

Of course, this is not the kind of detail you want. You want the meat! Figure 4.18 shows how a report for the most frequent searches might look.

Internal Search Report 1 [Data dissection]

This report is currently showing values for the URL parameter 's'.

	total	Jan	Feb	Mar	Apr
bounce rate	129	32	41	24	30
survey	02	10	15	20	29
dashboard	73	20	18	18	16
engagement	69	6	23	24	16
kpi	65	13	14	25	12
segmentation	65	10	17	17	19
seo	37	8	13	11	4
4q	35	5	8	14	8
trinity	35	11	13	7	4
reporting squirrel	33	4	1	1	27
None	305051	67095	01050	06477	67090

[More rows](#)

Figure 4.18 Report for internal site search keywords usage

This report gives an excellent peek into your visitor's mind and into what they are really looking for! You can take the report one step further by drilling down into a specific term, as shown in Figure 4.19.

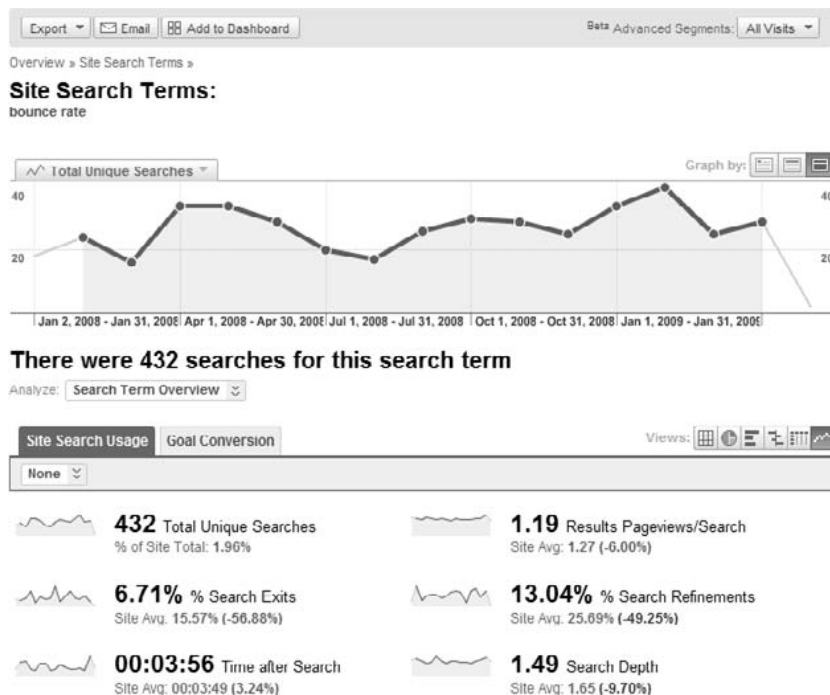


Figure 4.19 Comprehensive site search data for keyword searched by site Visitors: Bounce Rate

You can see the trend or interest and the seasonal effects and start to understand the performance of this particular query, whether by looking at the detailed statistics or by clicking the Goal Conversion tab.

It's time to kick it up a notch.

Measuring Site Search Quality

Now that you have established the importance of site search, you need to find out whether your site search engine delivers quality results.

Remember Bounce Rate? Of course you do. Some web analytics tools provide a *Bounce Rate for site search*. It is called % Search Exits. It measures the same phenomenon as Bounce Rate: the number of people who leave your website immediately after seeing the results provided by your internal site search engine. Figure 4.20 shows the report.

Search Term		Total Units	Individual Search Term performance: % Search Exits
1.	vecino	434	1.84%
2.	jilguero	390	30.69%
3.	lisa	369	1.36%
4.	linca	368	20.65%
5.	mila	359	3.06%
6.	gavilan	344	5.52%
7.	ad	343	71.43%
8.	gomez torres	326	3.38%
9.	alarcon	266	3.76%
10.	lobo	236	16.10%

Figure 4.20 Percentage of search exits. High = bad!

A quick glance at the bar graph on the right tells you which internal site search results are working and which ones are slacking off. In this case, queries 2, 4, and 7 (with a 71 percent exit rate!) need immediate attention. Perhaps your internal search engine algorithms are not right, or perhaps you don't have products, services, or pages relevant to them. Either way, it's cause for concern.

The other way to think about search quality is to measure the number of search results pages that are viewed by the visitor. We are all trained by major search engines to expect the most relevant result on top of the first page.

In measuring Results Page Views/Search, you can posit that if you provide the most relevant results on page 1, then Visitors will click one of the top listings and be on their merry way. Figure 4.21 illustrates how this metric looks in a report.

	Search Term	Total Uni	Individual Search Term performance: Results
1.	brain	44	1.23
2.	microscope	42	1.88
3.	drinking bird	39	1.15
4.	bella sara	37	1.60
5.	globe	34	1.26
6.	marble	34	1.41
7.	quercetti	31	1.10
8.	gears	26	1.92
9.	robot	26	1.35
10.	(not set)	24	1.71

Figure 4.21 Measuring search quality with Results Page Views/Search

The report shows the search queries where Visitors must dig deeper into your site search results to find what they are looking for. Consider the search query *gears*, with an average of 1.92 results views. It took two pages (on average 30 to 40 links to look through) to find what the visitor was looking for—that might be a bit much. It's time to look at what the results are on page 1 and fix 'em.



Note: Some analytics tools also automatically report on Time After Search (*time spent on your website after doing the search*) and Search Depth (*number of pages viewed after searching your site*). Both of these metrics might be interesting proxies for search quality, though it is hard to generalize them. For example, the visitor found what they wanted by using internal site search and bought the product right away (so, small Search Depth is better). Or the visitor found what they wanted on a content site and went on to read a lot more articles (so, large Search Depth is better). Use caution in how you end up using those metrics.

One final method of measuring search quality is, once again, to focus on customer behavior. Search Refinements help you understand how Visitors refine their queries to get optimal results.

Figure 4.22 shows the percent Search Refinements column in the Internal Site Search report.

	Search Term	None	Total Unique Searches	Results Pageviews/Search	% Search Exits	% Search Refinements
1.	bounce rate		39	1.33	2.56%	17.31%
2.	cross domain		31	1.00	83.87%	3.23%
3.	survey		25	1.36	0.00%	5.88%
4.	segmentation		16	1.94	18.75%	22.58%
5.	dashboard		15	1.73	6.67%	15.38%
6.	engagement		15	1.40	13.33%	14.29%

Figure 4.22 % Search Query Refinements for site searches

You can quickly see that Visitors refine their query for the word *cross domain* only 3.23 percent of the time, a likely indication that they find what they are looking for right away. But for the word *segmentation*, Visitors refine the query 19 percent of the time. That's not great. If your internal site search engine is good, then Visitors should not have to refine their queries. That is, they should not have to try searching again.

There is a small blessing in disguise here. With the results of Search Refinement, you can get a peek into your customers' heads. Figure 4.23 shows the Search Refinement report for the keyword *segmentation*. It is very helpful to know that most Visitors wanted to learn more about data collection when they were looking for information about segmentation. This information will help the company improve its site search results.

Site Search Refinement			
Refined Keyword	None	Search R	Search Refinements
1. data collection		9	1.57%
2. engagement		7	1.22%
3. bounce rate		4	0.70%
4. audit		4	0.70%
5. Excellent Analytics Tip		3	0.52%

Figure 4.23 Search refinement report for *segmentation*

Segmenting and Measuring Impact

By now you are not surprised by my obsession with segmenting data. Segmenting is always a valuable exercise. Most web analytics tools will allow you to do *inline segmentation*, or segment directly from the report itself. Inline segmentation is a very cool way for you to dig deeper and find insights, as you can see in Figure 4.24.

There were 7,164 unique searches via 3,977 search terms

Site Search Usage			
Total Unique Searches 7,164 % of Site Total: 100.00%	Results Pageviews/Search 1.27 Site Avg: 1.27 (0.00%)	% Search Exits 17.09% Site Avg: 17.09% (0.00%)	% Search Refinements 25.04% Site Avg: 25.04% (0.00%)
Search Term	Total Unique Searches	Results Pageviews/Search	% Search Exits
1. None	143	1.21	4.90%
2. Source	84	1.31	10.71%
3. Campaign	71	1.37	16.90%
4. Ad Content	69	1.41	7.25%
5. Visitor Type	66	1.33	16.67%
6. Landing Page	65	1.57	13.85%
7. Language	42	1.40	23.01%
8. Continent	36	1.34	5.26%
9. Sub Continent Region			
10. Country/Territory			
11. Region			
12. City			

Figure 4.24 Segmentation options for site search data

When you do this kind of segmenting, you should be able to answer a few questions easily:

- Do New Visitors to your site search more than Returning Visitors? For example, for my real estate website, Visitors from which city search more, and for what?
- Is there a difference in internal site searches done by Visitors from yahoo.com vs. those from google.com? Visitors who enter the site from my campaigns on custom landing pages should not be doing internal site search because I have created the most glorious and relevant landing pages, so are they still searching?

Finally, with segmenting, you can show the impact that site search has for the company's bottom line. When you start the process of analyzing this data, you might find that your site search stinks. So, you work hard to improve it to the extent you can. You want management to pony up for a new site search tool. They refuse. Are you stuck?

No. Show them, as with Figure 4.25, how Visitors who use site search end up converting at a higher rate than those that don't.

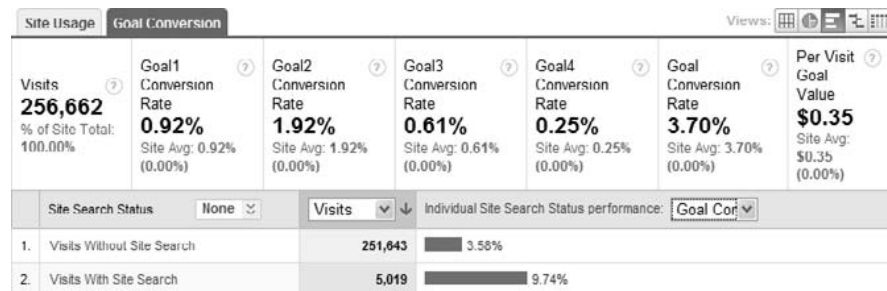


Figure 4.25 Goal Conversion Rate for Visits with and without site search

Although the number of Visitors who search is low, the Conversion Rate is multiple times higher. It is now easier to quantify the impact on Goals or Conversion Rate or Average Order Size if you invest in making site search better. You can measure the Conversion Rates, revenue, and so on, before you purchase the new tool.

That wasn't so hard, was it? You can do three simple and effective types of analysis on one of the most valuable sources of data in your possession.

There are related metrics and analyses you can do in your web analytics tool. For example, you can customize the JavaScript tag from your web analytics vendor and track the searches done on your site that did not yield any results back to your Visitors. This information can be important, especially if your site search is really poor. Typically, though, these kinds of searches are done by a very small number of your website Visitors. Therefore, you should first do the three types of analysis I have covered in this section. Eat all the yummy low-hanging fruit first; you'll have a larger impact on most of your website Visitors. Then move on to things that are important but will have a smaller overall impact.

Search Engine Optimization (SEO) Analysis

Search engines are a key part of any company's acquisition portfolio. For good reason, citizens of the Internet use search to find everything they are looking for, including you. The business impact of this customer behavior is that we all pour a lot of resources into optimizing our websites to show up optimally in search engines for relevant queries.

Figure 4.26 shows the search engine results page on www.google.com for a typical query.

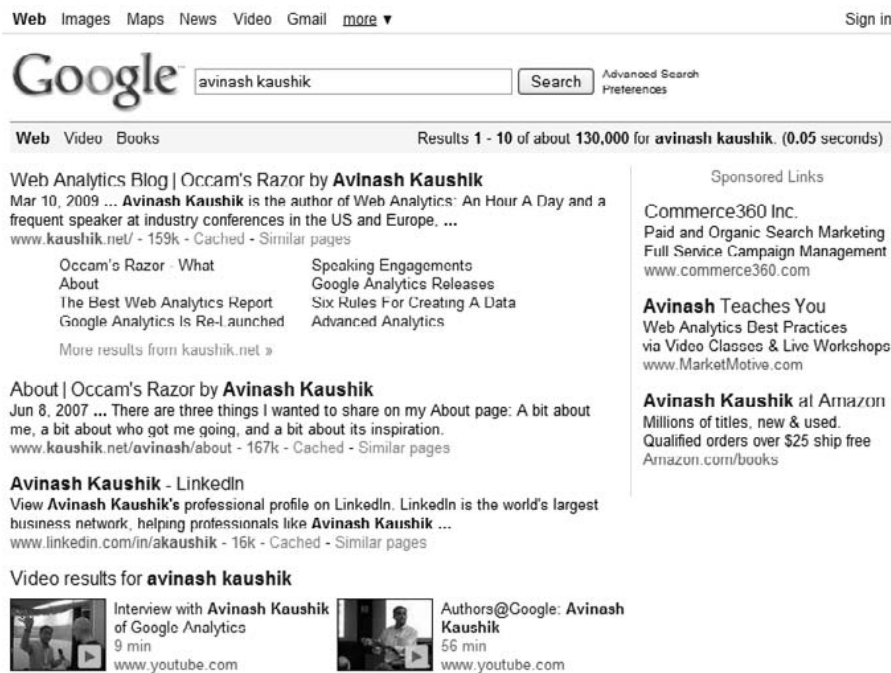


Figure 4.26 Search results page on Google

Sponsored links are the paid search results (I'll cover how to analyze these in the next section). The search results in the main body are the organic search results, or those that the search engines list for free based on their unique algorithm.

You can improve your organic results by taking a number of actions on your website. These include having a website that is easily indexable by the automated robots from the search engines, ensuring your URL structure is clean, using JavaScript in links judiciously (robots don't execute JavaScript and hence won't follow those links), ensuring that the content on your pages is relevant and uses the right keywords, and so on.



Note: To learn more about SEO strategies, you can download the official Search Engine Optimization Starter Guide from Google (<http://sn.im/googleseo>). You can also purchase the wonderful book *Search Engine Optimization: An Hour a Day* by Jennifer Grappone and Grativa Couzin (<http://sn.im/akseohour>). For the technically inclined amongst you, the CD accompanying this book includes “The Web Developer’s SEO Cheat Sheet” from SEOmoz.

SEO analysis is unique because the data partly exists in your web analytics tool (such as Omniture, Webtrends, or Affinium NetInsight) and partly in external tools such as Webmaster Tools provided by Google and Microsoft.

In the next section, I’ll cover analysis that will measure four facets of your SEO strategy: current performance (traffic), content coverage (indexing by search engines), keyword performance (search engine results), and Outcomes.

Performance: Traffic from Organic Search

You always want to know whether you should be engaged in this whole endeavor at all! Luckily, every web analytics tool provides quick access to understanding organic search traffic trends, as shown in Figure 4.27.

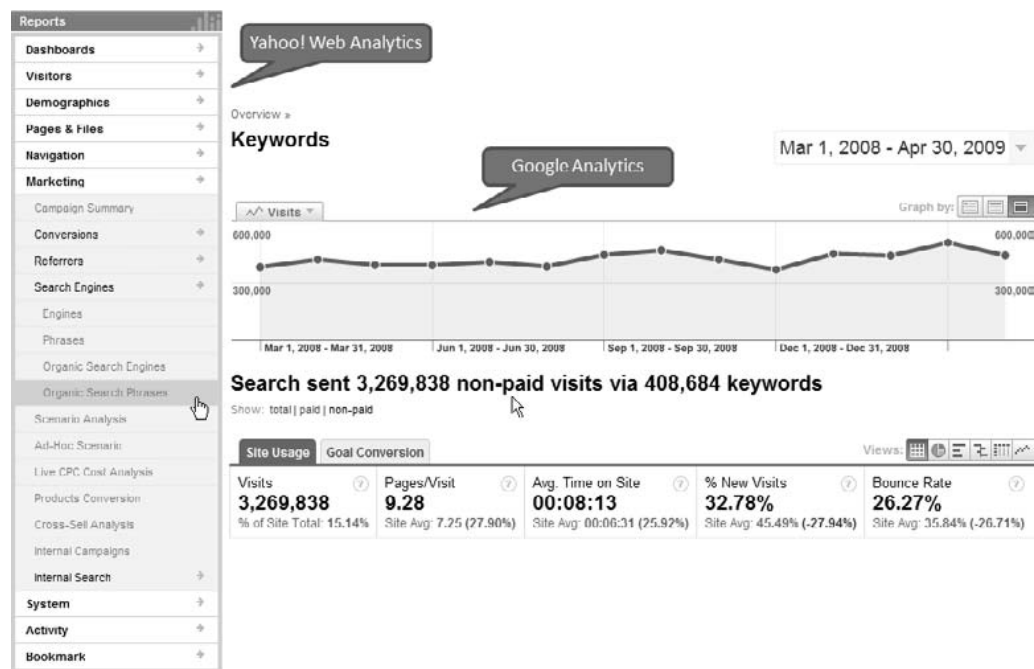


Figure 4.27 Organic search traffic reports

At a glance, you can see key statistics as well as understand the performance of the organic search traffic by looking at metrics such as Bounce Rate (looks fabulous here!).

You can also click the Goal Conversion tab and quickly assess the value of organic traffic. Are they converting at an optimal rate? Higher? Lower?

But the truly insightful bit is to segment the overall traffic trend and understand the performance of paid search and organic search as a piece of that bigger macro puzzle. Figure 4.28 shows the report you need to create immediately. It includes a 13-month trend of overall traffic (All Visits), Paid Traffic, and organic traffic (Non-paid Traffic).

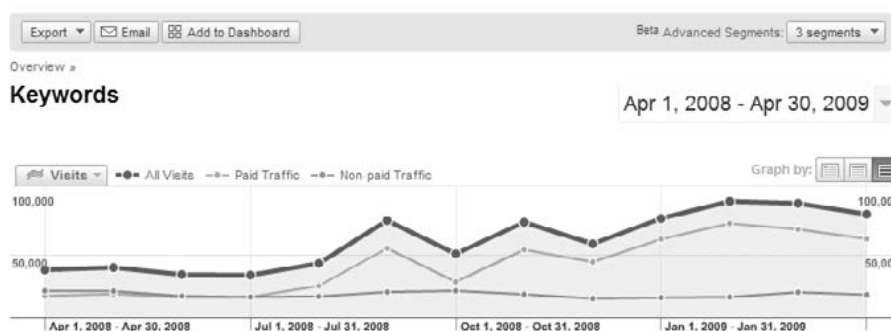


Figure 4.28 Segmented long-term paid and organic search report

From April 2008 through July 2008, organic and paid traffic contributed about the same amount to the overall traffic. Then something started to shift, and with almost every passing month (except October), the paid component kept growing and becoming increasingly more important to the business. Organic search, on the other hand, essentially remained flat.

Typically this type of trend should cause you to dig deeper. Why did the paid search strategy get increasingly successful while the organic search strategy went nowhere? What is unique about our website and content? Is our business (content, products, services) so dynamic that organic search won't work for us? Why the "over-reliance" on paid search?

An effective search strategy is a portfolio strategy. You must optimize for all major search engines, and you must effectively use paid and organic search. Doing only paid or only organic is suboptimal. Optimizing your organic search strategy starts with reports like those shown in Figure 4.27 and Figure 4.28.

Content Coverage: Indexing by Search Engines

With your organic search strategy, your explicit goal is to get your website indexed properly by the search engine. If this happens, then you increase the chances that your site will actually show up when people use search engines to look for *stuff* relevant to you.

You can measure the impact of your SEO efforts in terms of content coverage in two ways:

- You can measure the amount of content being indexed over time (this should go up if you publish new content).
- You can measure the number of pages on your site that get traffic from search engines.

All three major U.S. search engines provide a service affectionately called Webmaster Tools. This is your first stop in understanding whether your site is primed for search success.

Before you can use the tool, you or perhaps an IT person at your company will have to sign up with the search engine and authenticate yourself by placing a small XML or HTML file in your website's root folder (hence proving you own the site and can see the data).

In Google, you review the Crawl Stats report, as shown in Figure 4.29.

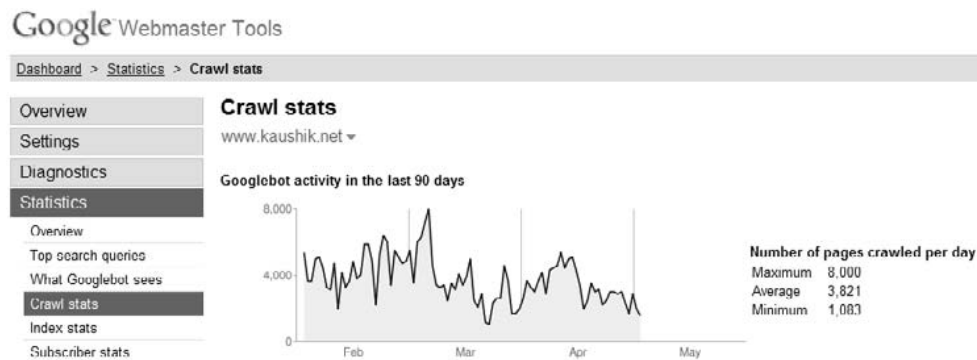


Figure 4.29 Google Webmaster Tools' Crawl Stats report

Contrary to expectations, as you publish more content on the website and you expend major SEO efforts, fewer pages were indexed in the last 45 days. You can't find this data in your web analytics tool. It is important to keep a close eye on it.

Figure 4.30 shows the statistics from Yahoo! Site Explorer.



Figure 4.30 Yahoo! Site Explorer's Statistics report

Same data, different numbers. It is confusing, but each search engine crawls and categorizes pages on your site in a different manner. Yahoo! shows 14,178, and Google shows around 3,821. Don't get stressed about the actual number; just compare the trends over time. The rule is simple: if you keep adding content to your site, the numbers should go up if the search engine is crawling your site optimally.

If your IT team is giving you a tough time, you can also try to use other tools to get at this information. Marketleap provides two useful reports at <http://sn.im/mktleap>. The Search Engine Saturation report shows you the number of pages a given search engine has in its index for your site, and the Link Popularity Check report shows the total number of links that a search engine has found for your site. The nice thing about the Marketleap reports is that you can trend the data over time.

Having checked the search engine perspective, you now need to check the view from your website's perspective. The content coverage report that you want to run is # of Landing Pages for Organic Traffic (meaning the entry pages on your site that get organic traffic from search engines, not the custom pages you created for paid or email campaigns).

Figure 4.31 shows how the report looks in Google Analytics; your web analytics tools should easily provide this precise view.

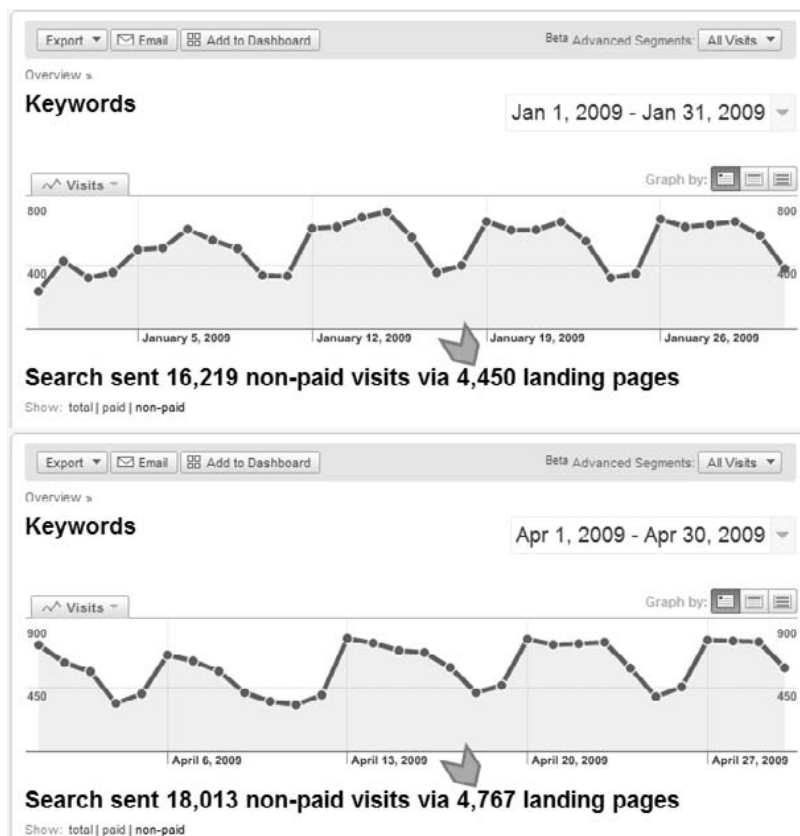


Figure 4.31 Content coverage report: organic search landing pages

The hypothesis is a very simple one. If your website is being indexed correctly and your SEO efforts are working, then over time the number of pages that get direct traffic from search engines will increase. In Figure 4.31 you can see how over a three-month period of exhaustive and often painful SEO work, both the number of Visits to the website as well as the # of Landing Pages for Organic Traffic has increased at a nice clip.

Keyword Performance: Search Engine Results

Search engines have increasingly moved beyond showing text results to including images, videos, and other relevant kinds of content. This concept is often called *universal search*. Additionally, search engines often customize results based on a whole host of factors such as your location, past search queries, language, and so on.

Therefore, you need to look beyond the keywords in your analytics reports and try to understand whether you are *showing up* for the optimal set of keywords in search engine results. And you want to know the surprises—unexpected keywords that show up in results.

An excellent way to measure these surprise keywords is to use the Webmaster Tools report. Figure 4.32 shows the Top Search Queries report from Google's Webmaster Tools (<http://sn.im/gogwmt>).

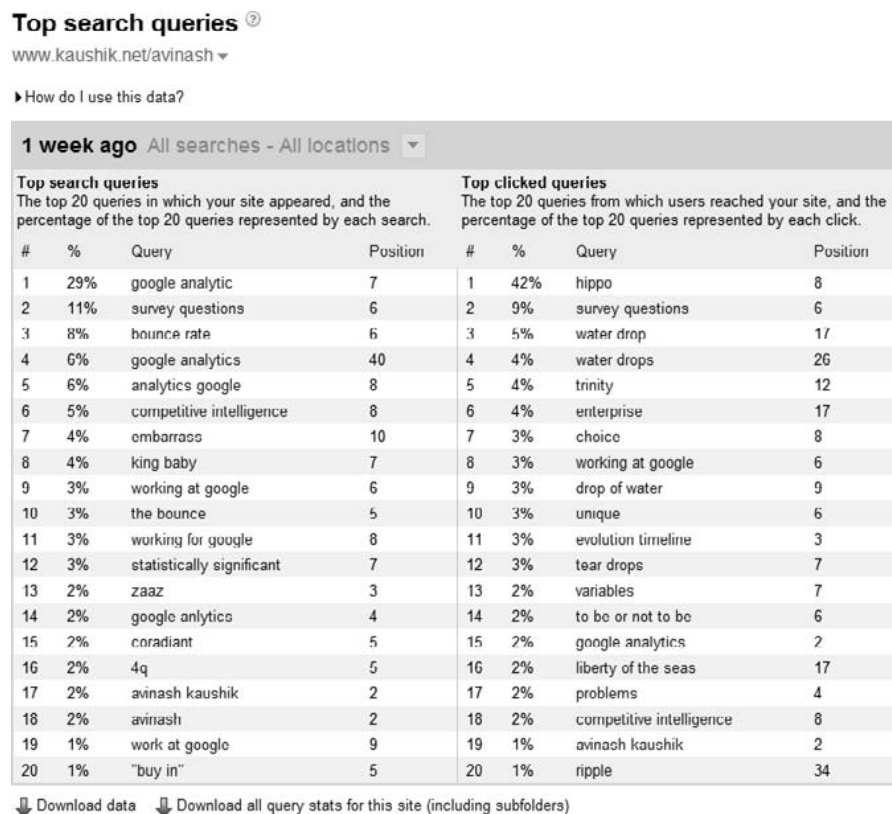


Figure 4.32 Top Search Queries report, Impressions vs. Traffic

On the left are Impressions, or user queries for which your website appeared in search results. On the right, Traffic shows queries where users actually clicked your website's link on the search results page and came to your website. As you might have guessed, the Traffic data will be in your web analytics tools, but your tools are completely blind to the Impression data because it exists only in the search engine. That is what makes the Impression data particularly valuable.

When you analyze the data, look for surprises. Your site shows up for a whole bunch of relevant results (1, 2, 3, 4) and less than relevant results (7, 8, 20). You can optimize your robots.txt file or your website content to ensure you are showing up for all the right keywords.

At the same time, not all keywords for which you have impressions (1, for example) result in traffic to you (or does not exist in Traffic on the left...sad). In this case, you should go back and look at the page that shows up—whether the page has the right snippet, whether the title of the page is right, and so on. Such changes will help make the page stand out again and yield more traffic to you.

The actionable insights from this data will feed into your SEO program. As your team implements these keyword-level strategies, you can measure their success by using the monthly data available from the search engine (see Figure 4.33).

December				April			
Top search queries				Top search queries			
The top 20 queries in which your site appeared, and the percentage of the top 20 queries represented by each search.				The top 20 queries in which your site appeared, and the percentage of the top 20 queries represented by each search.			
#	%	Query	Position	#	%	Query	Position
1	24%	google analytic	9	1	18%	google analytic	9
2	19%	google analytics	39	2	13%	survey questions	10
3	9%	survey questions	8	3	8%	bounce rate	6
4	9%	analytics google	8	4	7%	google analytics	45
5	6%	bounce rate	3	5	7%	the bounce	5
6	4%	damini	7	6	7%	analytics google	9
7	3%	google	397	7	5%	competitive intelligence	10
8	3%	analytics	45	8	5%	working at google	7
9	3%	competitive intelligence	8	9	5%	king baby	6
10	2%	butt	54	10	4%	working for google	7
11	2%	avinash	3	11	3%	statistically significant	5
12	2%	zaaz	3	12	3%	avinash	2
13	2%	trinity tech talk	9	13	2%	next stop wonderland	9
14	2%	www "google be"	6	14	2%	survey examples	6
15	2%	metric	6	15	2%	zaaz	4
16	2%	occam's razor	15	16	2%	avinash kaushik	2
17	2%	google analytics	4	17	2%	work at google	9
18	1%	razor	45	18	2%	work for google	7
19	1%	avinash kaushik	2	19	2%	coradiant	8
20	1%	what is bounce rate	8	20	2%	4q	7

Figure 4.33 Comparing improvement in Impressions over time

It is clear that the implemented SEO improvements have yielded positive results when you compare the Impressions data for December and April. For example, the

phrase *statistically significant* was specifically targeted for optimization. It was non-existent on the left (December) but shows up at #11 in the report on the right (April). Hurray!

Outcomes: Goals, Revenue, and ROI

In the SEO business, we love our rankings, engines, and results, but we rarely give that same love to the Outcomes we are driving for our companies. We seem to be obsessed with playing with the search engines.

Consider this last point to be my plea to focus on measuring Outcomes, with the same vehemence you bring to your email, affiliate, or paid search campaigns. Doing so will ensure that you get the funding you need to execute your SEO efforts and show the value of this free traffic stream.

You can measure a whole host of ecommerce and non-ecommerce Outcomes. I'll recommend at least two.

First, measure the impact on your business of the traffic that comes from organic search (from all search engines), as shown in Figure 4.34.



Figure 4.34 Segmented conversions: All Visits vs. organic search traffic (Non-paid Traffic)

Although organic traffic makes for a small percent of the website's traffic, it actually accounts for an outsized percentage of its multiple conversions. If you know a better way to get an immediate bonus as an SEO professional, then let me know.

If you have an ecommerce website, then you can dive even deeper into the data and analyze key performance indicators such as Revenue, Average Order Value, Products Sold, and so on, as shown in Figure 4.35.

Analyzing this data in context with your other campaign data will help you measure true ROI (remember to also add the costs of doing your email, affiliate, or paid search campaigns in computing ROI).

I'll give you one final example to illustrate the coolness of measuring Outcomes for organic search. Say you get a ton of traffic. You do your normal analysis and tell your boss that most of the traffic is irrelevant and you want to eliminate it. Your boss will faint. Get some cold water, and sprinkle it on her face. After she gets her wits

together, tell her that the reduced traffic won't have any impact on Conversions; in fact, you can improve them.



Figure 4.35 Ecommerce metrics for organic traffic

The sweet delight is that you can easily measure the result of the shifts you are suggesting in your organic search strategy (see Figure 4.36).

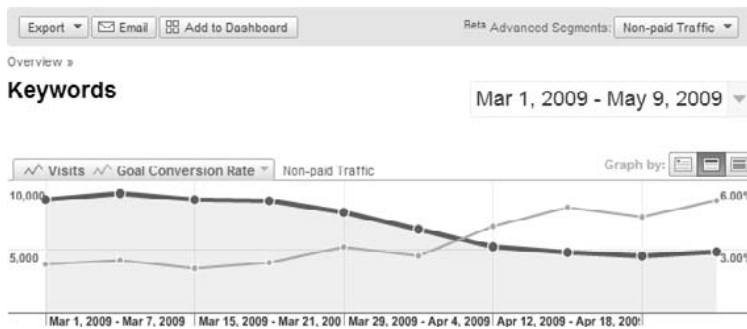


Figure 4.36 Correlating organic search traffic with Goal Conversion Rates

You can pinpoint when you started to implement your new and improved search strategy by focusing on the right keywords and web pages (March). Sure enough, traffic starts to go down, and mercifully Conversions pick up slightly. You have some sleepless nights, but your data-driven SEO strategy starts to clearly show results by the end of March (traffic is down almost 50 percent, yikes!). Of course, you are focusing on the line that's going up, and that's Conversion Rate).

As analysts, and as marketers, the strategies we follow and the recommendations we make might seem counterintuitive. But if you are wise and tie your recommendations to measurable Outcomes, you can get away with almost anything.

Pay Per Click/Paid Search Analysis

Paid and organic search are conjoined twins. You can't have one without the other. Any company that does both effectively (and holistically) stands to reap huge rewards.

Since there are entire books written about PPC analysis, in this section I hope to share with you the starting points of effective analysis of your PPC campaigns.

The very first thing you must do is *teach* your web analytics tool which click from the search engine is from an organic listing and which is from a pay-per-click (PPC) listing. Analytics tools (with the exception of Google Analytics/AdWords) don't come with native intelligence for separating organic and paid clicks. Each search engine is unique and sends a referral string that is distinct. Furthermore, each web analytics tool has its own way of storing search referral string data, which means if you put the wrong tracking parameter in the wrong place, you are hosed.

Let's look at an example. If you search the word *Omniture* on www.google.com today, you get two listings—one organic and one paid. The organic listing goes to the URL www.omniture.com. The paid listing goes to the URL www.omniture.com/static/1923?s_scid=TC|5379|omniture||S|e|2831181395.

Everything you see after the question mark (?) in the previous URL is known as the *referral string*, and it contains a tracking parameter that tells Omniture which click came organically and which came via PPC. Without the tracking parameters, both clicks would be considered as coming from organic search.

If you had done a search on www.yahoo.com, you would also get an organic listing and a paid listing for Omniture. The paid listing goes to the URL www.omniture.com/static/278?s_scid=680217600000000309&clicksource=standard&OVRAW=omniture&OVKEY=omniture&OVMT=standard&OVADID=4822371011&OVKID=130976483511.

You can see how the referral string is organized differently on Yahoo!

I cannot stress how important it is that you make sure that all your search campaigns are tagged properly and according to the rules imposed on you by your web analytics tool, or else your ability to do analysis will be kaput.

The only exception to this rule is if you use Google Analytics as your analytics tool and use Google AdWords to do your PPC campaigns. In this case, simply tying your AdWords account to your Google Analytics account will ensure that you don't have to tag your campaigns in AdWords to teach Google Analytics anything; Google does everything for you.

But when you analyze your Microsoft or Yahoo! paid search campaigns in your Google Analytics account, you must tag them properly (using the free URL Builder tool; <http://sn.im/urltag>).

The following are the kinds of techniques that you'll use to optimally analyze your PPC campaigns.

Performance: Traffic from Paid Search

Some of the analysis you'll do for your PPC campaigns is similar to your organic search campaigns. For example, you will first impress your co-workers with the analysis shown in Figure 4.27 and Figure 4.28. The goal is to highlight the long-term trends for PPC campaigns. Are you getting better? Worse? Identify the major shifts in traffic, and identify the rationale.

The business impact analysis you'll perform for your PPC campaigns is also very similar to what I have already covered, as you saw in Figure 4.34, Figure 4.35, and Figure 4.36. The goal is to highlight how PPC campaigns add to the bottom line. Do you see the same kind of efficiency in your paid campaigns as you see for your organic campaigns in Figure 4.36? If not, why not?

Each of the following five analyses will help you understand how important your PPC campaigns are and help you understand priorities and initial focus areas.

Measuring the End-to-End View

The starting point of your PPC analysis will look like Figure 4.37.

AdWords sent 830 visits via 10 ad contents

Site Usage

Goal Conversion

Ecommerce

Clicks

Views:

Visits

830

% of Site Total: 0.97%

Pages/Visit

5.40

Site Avg: 3.91 (37.96%)

Avg. Time on Site

00:03:13

Site Avg: 00:02:14 (43.96%)

% New Visits

92.77%

Site Avg: 90.72% (2.76%)

Bounce Rate

18.80%

Site Avg: 53.44% (-64.83%)

	Ad Content	None	Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	Icky Toys		209	6.41	00:03:41	96.17%	28.71%
2.	C3000 Chemistry Kit Sale		191	4.99	00:03:05	92.67%	14.14%
3.	Marshmallow Shooter Fun		134	4.66	00:02:25	90.30%	17.91%
4.	C1000 Chemistry Kit Sale		121	4.66	00:03:07	91.74%	11.57%
5.	C2000 Chemistry Kit Sale		50	4.20	00:02:51	86.00%	14.00%
6.	Perfume Science Kit Sale		50	4.46	00:02:20	94.00%	16.00%
7.	Give Smiles & Smarts		38	6.89	00:04:33	92.11%	23.68%
8.	Who gave you that toy?		37	8.30	00:04:57	94.59%	18.92%
9.	Smarter Smiling Toddlers		0	0.00	00:00:00	0.00%	0.00%
10.	Trainmech at BrainWaves		0	0.00	00:00:00	0.00%	0.00%

Find Ad Content:

containing

Go

Go to:

1

Show rows:

10

1 - 10 of 10

Figure 4.37 Clickstream analysis of paid search campaigns





You can quickly assess which keywords (or ad groups or user queries) bring clicks to your site and how each performs in terms of initial metrics (Time on Site, % New Visits, and so on). But look at the tabs in Figure 4.37, and you'll realize you can go deeper into the data. So in Figure 4.38, you can measure Goal Conversions, e-commerce, or, my favorite, the end-to-end view (Clicks).


Site Usage

Goal Conversion

Ecommerce


Clicks

Views:    

Visits 


830

% of Site Total: 0.97%

Impressions 


253,170

% of Site Total: 100.00%

Clicks 


934

% of Site Total: 100.00%

Cost 


\$551.53

% of Site Total: 100.00%

CTR 


0.37%

Site Avg: 0.37% (0.00%)

CPC 


\$0.59

Site Avg: \$0.59 (0.00%)

RPC 


\$1.07

Site Avg: \$32.53 (-96.71%)

ROI 

81.02%

Site Avg: 5,408.41% (-98.50%)

Margin 

44.76%

Site Avg: 98.18% (-54.41%)



	Keyword 	Visits 	Impressions	Clicks	Cost	CTR	CPC	RPC	ROI	Margin
1.	educational toys	267	234,110	275	\$172.99	0.12%	\$0.63	\$0.03	31.61%	24.02%
2.	marshmallow shooter	95	3,434	117	\$59.97	3.41%	\$0.51	\$2.02	294.40%	74.64%
3.	chemistry kits	70	468	80	\$46.45	17.09%	\$0.61	\$0.00	100.00%	0.00%
4.	chem c1000	57	354	59	\$44.66	16.67%	\$0.76	\$7.53	895.16%	89.95%
5.	chemistry kit	55	624	63	\$37.57	10.10%	\$0.60	\$0.00	100.00%	0.00%
6.	perfume science	43	855	50	\$29.34	5.85%	\$0.59	\$1.08	83.88%	45.62%
7.	chem 1000	42	510	54	\$30.58	10.59%	\$0.57	\$0.00	100.00%	0.00%
8.	chem c3000	34	207	33	\$23.17	11.50%	\$0.70	\$0.00	100.00%	0.00%
9.	chem c2000	29	216	35	\$16.47	16.20%	\$0.47	\$0.00	100.00%	0.00%
10.	chem 3000	20	269	26	\$8.94	9.67%	\$0.34	\$0.00	100.00%	0.00%

Figure 4.38 Analyzing the end-to-end view of success

This lovely report manages to bring three different pieces of data together so you can judge effectiveness in one nice view. The first set of data, the middle box, is data from the search engine (in other words, AdWords). It shows the performance of your advertising via the number of Impressions, Clicks, Cost, CTR (which is the click-through rate), and CPC (which is the cost per click). You'll understand what it takes to get someone to your site by individual keywords. Check out the first row, "educational toys," with 234,118 impressions to get 275 clicks! Oh my.

The second set of data should be very familiar: Visits. Remember, a click is not the same as a Visit. During one session, a Visitor could have clicked three different ads from the search engine. That would show up in your reports as Clicks = 3, Visit = 1.

The last piece of data, on the right, illustrates the business impact metrics: RPC (the revenue per click), ROI (the return on investment), and Margin (typically revenue minus campaign cost). This is a great way to understand if you had a positive or negative impact on your business.

One caution, though. Notice I said "campaign cost" when computing Margin. By default, almost all web analytics tools will not include the cost of goods sold (COGS), that is, what it costs you to make the product or service you are selling, in the Margin calculation. Hence, you need to take that number with a grain of salt. If your web analytics tool allows you to import the COGS data, then do that and compute your true Margin. Omniture, Webtrends, Coremetrics, and many other tools allow you to do this.

Looking at your end-to-end view is important because you want more than clicks to your site. You want more than Visits. You want bottom-line magnifying impact. This simple report is very effective at showing you that impact.

Analyzing PPC Ad Position

So, here is how PPC campaigns work. You pay. Your bid goes into an auction. Others bid. They compete with you. A smart algorithm gets applied. Your ad gets listed on the search engine results page based on the algorithm + competition + bid. Now the cute part is that the position of your ad will also depend on those three factors (amongst others).

Hence, understanding the impact of different ad positions can be an insightful set of analysis you can perform. Figure 4.39 shows one such report, Keyword Positions, from Google Analytics. In other web analytics tools, you'll also easily find Average Position reports.



Figure 4.39 Keyword Positions report

The Keyword Positions report is always chockful of insights. Each box on the bottom right shows the number of clicks your ad received when it was listed in a certain position. You quickly realize that you pay a different amount of money to the search engine depending on the ad position—typically, the higher the position (Top 1, 2, 3), the more expensive.

Notice in Figure 4.39 that the keyword *trabajo* has an interesting distribution of clicks to your site. You get more clicks in position Side 1 than you do from Top 2 or Top 3. That seems counter to what you might believe (“Got to be in Top 1, 2, or 3 to get any clicks!”).

So, your first insight is realizing which position to bid for your top keywords to ensure the optimal number of clicks. But, who cares about clicks? The Web Analytics 2.0 ninja cares about adding business value!

Besides displaying multiple dimensions of data in a way that an Excel table never can, the Keyword Positions report gives you that drop-down menu right above the search engine logo, Position Breakdown. Here you can choose the metric that is most applicable to your business, such as Pages/Visit, Time on Site, Bounce Rate, various Goal Conversion Rates, Per Visit Goal Value, and so on.

For our example, the company is a content-only website, with no ecommerce, and its goal is not simply to get a Visit but rather to get Visitors who spend a lot of time on the website. They'll choose Avg. Time on Site from the drop-down menu and see the report shown in Figure 4.40.

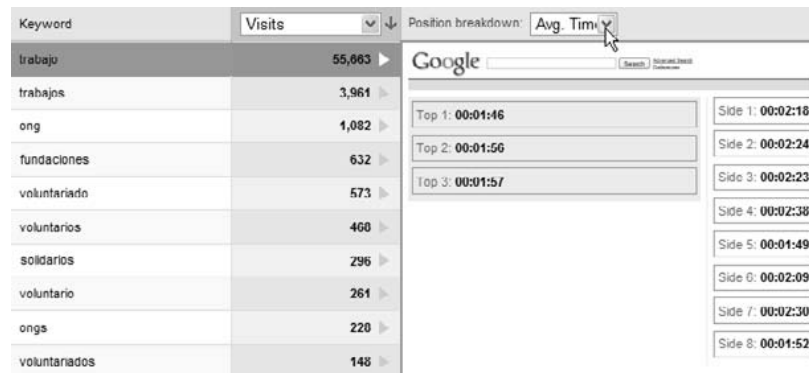


Figure 4.40 Keyword Positions report for Avg. Time on Site

You should notice two things. First, the report shows positions data from the search engine (AdWords in this case) with the metric for Time, which is from your web analytics tool. Without you realizing it, the report is making some pretty nifty connections. Second, you'll immediately notice that the *ego positions* (Top 1, 2, 3) are doing OK, but the top four positions on the side (cheaper bids!) actually deliver Visitors who spend more time on your website.

With these insights, you can balance between the number of clicks you can get from each position and your business success (time). You can tweak your bids to ensure you get the best position at the best cost.

Measuring PPC Customer Behavior

At the beginning of this chapter, I covered the importance of focusing on customer behavior rather than aggregate metrics. You can do some cool PPC analysis by applying that principle.

The setup is quite simple. Let's measure how long it takes for someone to convert from the first time they visit your website. The report is called Days to Purchase, which is shown in Figure 4.41. You'll notice, top right, that it is segmented to show only Paid Traffic.

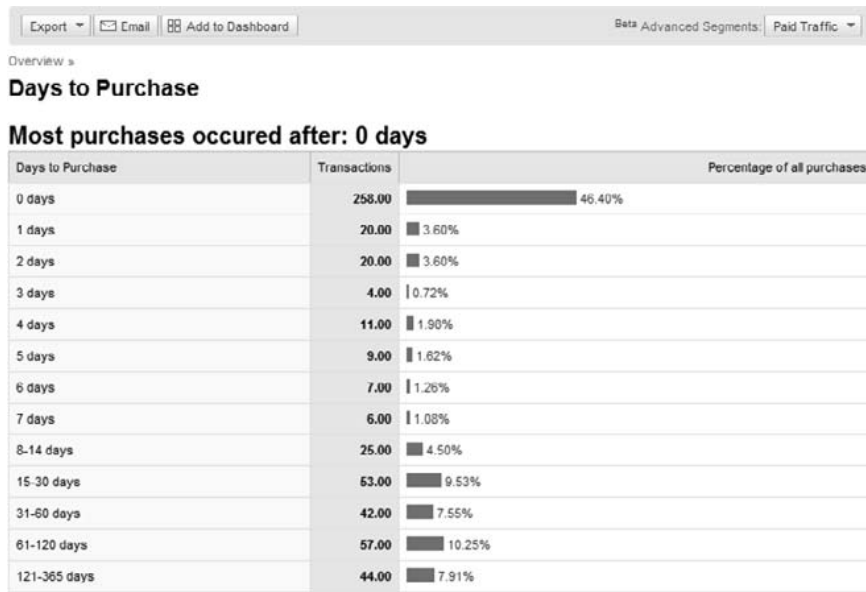


Figure 4.41 Days to Purchase, PPC campaign traffic

This data is for a travel website. It might strike you as odd that only 46 percent of the people make a purchase on the same day, simply because airline tickets and hotels and cruises all tend to get more expensive with each passing day. That's odd behavior for Visitors.

OK, let's find out whether we see the same behavior with traffic from other sources. Figure 4.42 shows the report for Direct Traffic.

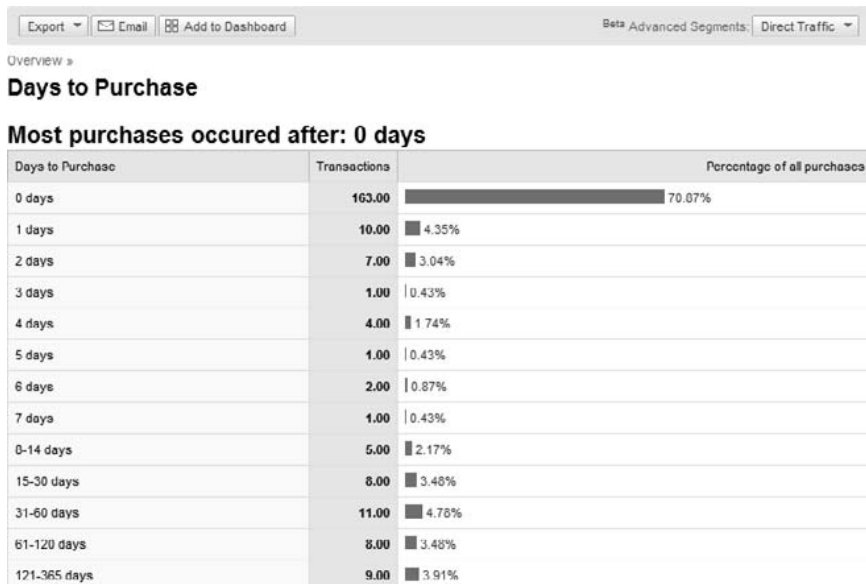


Figure 4.42 Days to Purchase, Direct Traffic

Whoa! With Direct Traffic, 71 percent purchase on the same day. That is normal behavior for Visitors for a travel site. So, what is up with the PPC traffic? Why don't they behave like *normal people*? If the behavior from PPC traffic is so odd, then how can we treat them better or differently?

These are all great questions, and they're the reason why I recommend this type of analysis. From my stress on segmentation, you already know that each traffic stream is unique.

In this case, the travel website took this data and reviewed the landing pages for its PPC campaigns. At the time of the previous report, it was geared toward converting Visitors quickly ("Buy now! Book now! Give us your money, now!").

The first action the marketers took was to soften the calls to action, because they realized a good chunk of the traffic does not want to buy right away. Then the marketers added a new feature, Save Your Itinerary. They realized Visitors would come back, so they might as well make it easy for them.

Finally, the marketers added another feature: "Email me if price goes up by x percent," where x was a number that the Visitors could input when saving an itinerary. This was a very clever move because the travel agency then had the contact information for the Visitors and could email them when the price went up by 10 percent or 20 percent or whatever number the customer input. This caused the customers to return and make a purchase sooner, and the customers were happier because they felt the site was watching out for them.

The net return for closely analyzing PPC customer behavior was that it brought forward Conversions to fewer days to purchase and tripled Conversion Rates. That's not too shabby for a simple segmented report, right?

There are lots of different kinds of analysis you'll do for your PPC campaigns. My hope is that these four *non-normal* examples—measuring traffic performance, measuring the end-to-end view, analyzing ad position, and measuring customer behavior—highlight the kinds of analysis that will help you maximize profitability. Advanced paid search analysis techniques are discussed in more detail in Chapter 11.

Direct Traffic Analysis

I am a huge fan of Direct Traffic because it is free. Who does not love free? Yet analysts and marketers constantly get enamored by campaigns and search and other perceived cool things and ignore this valuable source of traffic.

Direct Traffic is typically defined as *noncampaign, nonsearch, nonlinked* traffic, that is, Visitors who come to your website by typing in your website's URL or who come through a bookmark.

Direct Traffic is valuable because these Visitors know you already, they tend to be your existing customers, and you are not paying a campaign *bounty* to get them to come to your site. It pains me that some web analytics tools don't even have a clearly identified bucket for this traffic stream.

Capturing Direct Traffic

To accurately capture Direct Traffic, you need to do the following:

- Tag all your website pages so you won't lose the referring/campaign information the Visitors might bring with them.
- Make sure your different campaigns contain the proper tracking parameters.

If you forget to do either of these tasks, then your campaign, search, and referral traffic will be wrongly classified as Direct Traffic.

I highly recommend that you make a key part of your analytics strategy understanding free traffic and identifying strategies to nurture and grow this segment of your Visitors. You'll make your loyal customers happier, and you'll end up saving on acquisition costs. Do I hear a win-win?

The kinds of analysis you'll do for your Direct Traffic will be quite similar to what you do for your other traffic streams. Let me share a few reports that I use when I start with my analysis.

Reviewing Current Performance

First you want to educate yourself and your decision makers about the current state of affairs. That is, you want to know how you are doing today in terms of the performance of Direct Traffic and whether there are opportunities to grow. Figure 4.43 shows how the data appears in your web analytics tool, in this case Google Analytics and XiTi.

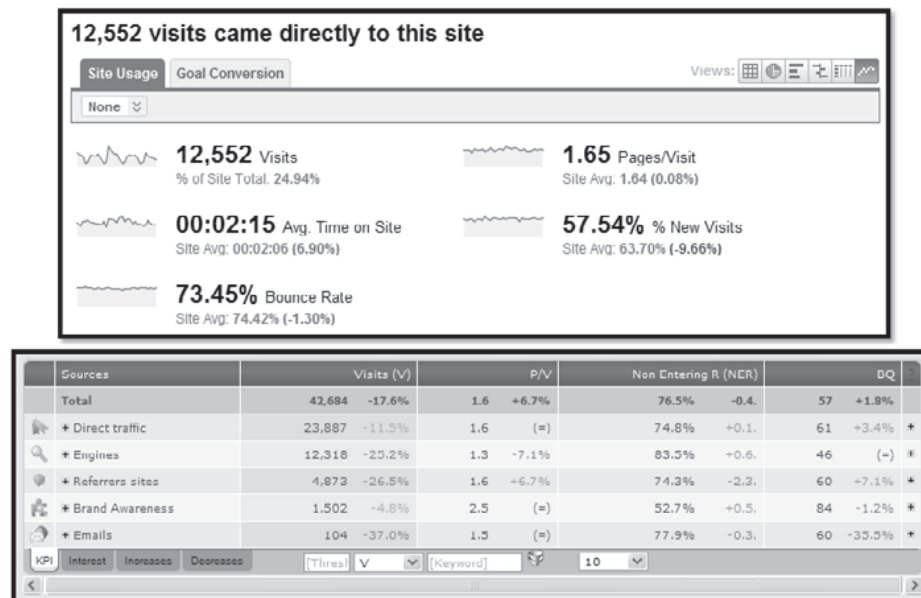


Figure 4.43 Direct Traffic reports for Google Analytics (top) and XiTi (bottom)

In this type of report, you need to understand what percentage of traffic is Direct Traffic. You can clearly see that percentage in the top part of Figure 4.43: 12,552 Visits form a total of 24.94 percent of website traffic. Not too bad, but it could be better. In the bottom, the Direct Traffic is 56 percent of the site traffic.

Next you need to dive into performance. Look at the base metrics such as Time on Site, Bounce Rate, Pages/Visit, and so on. It is very clear in the Google Analytics report (top) in Figure 4.43 that Direct Traffic performs very well when compared to Site Averages for those metrics.

An unsurprising close to this section is the recommendation to review the longer-term trends for Direct Traffic. Over the past 12 months, is the line graph up and to the right? If not, then you need to find out why not. How about when you compare it to other segments of traffic on your site?

Understanding the Opportunity and Educating Management

We will dive deeper into understanding Outcomes in the next chapter, but I want to stress here that you never convince decision makers to take you seriously without showing impact. Figure 4.44 shows a report that will be of incredible value in your effort.

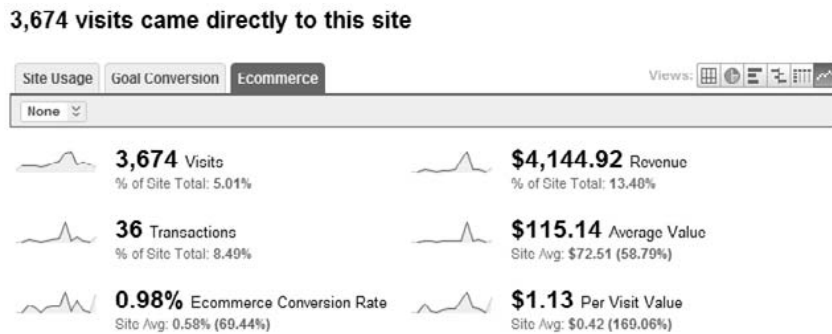


Figure 4.44 Ecommerce report for Direct Traffic

For your website, the impact might not look exactly like Figure 4.44, but it is not unusual at all for Direct Traffic to perform at a higher level than your other traffic sources. The reason is simple; these are people who found you without you having to incentivize them to come. They know you already, and usually they are going to deliver superior Outcomes.

Take both steps discussed earlier: identify the performance of the core metrics (as in 36 Transactions, \$4,144 in Revenue, and so on), and clearly highlight that performance (as in the comparison to Site Total). The latter is cool because comparing the metric to the Site Total helps highlight value and gains you some love. For example, you'll notice that Average Value from Conversions of Direct Traffic is 58.79 percent higher when compared to the Site Average. That is rather sweet, and stuff like that helps you get support from your lovely management team.

Segmenting to Understand Uniqueness

Given the unique nature of the people who come directly to your site, it is extremely productive to understand them. It's like watching a pride of lions and taking notes. What is unique about them? What do they like to eat; that is, what content do they like to consume? Why do they return to the same place? You catch my drift.

The following are my favorite kinds of analysis:

Content analysis Note what pages/content directories these people visit most often. Look at their click pattern in the site overlay (click density) report—what does it indicate in terms of their preferences? Note whether the internal site search data indicates particular needs in terms of products, services, or other needs. For example, analyzing Visitor Loyalty and Depth of Visit could reveal that most Direct Traffic visits frequently and tends to read only about sports and entertainment, while all other traffic sources seem to gravitate toward politics and culture. Actionable insights!

Purchase behavior Visitors who come through Direct Traffic are often, but not always, existing customers, and understanding what they buy is an insightful exercise. Or in case of non-ecommerce websites, you want to understand the behavior around submitting leads or contributing donations and other such Outcomes. This data can be especially useful for future targeting possibilities; this is a delight if you have analyzed the behavior to understand seasonal impact.

For example, a business may recognize that in the first three weeks after releasing the latest version of a product, only existing customers made the purchase—with no need for any campaigns, coupons, or other enticements. Then came the new customers. You would execute your merchandizing and content strategy on the site completely differently if you understood this.

Direct Traffic is valuable. It offers an opportunity for you to maximize the potential of a relationship you already have for a source of traffic that does not typically cost you much, if anything, in terms of acquisition.

Email Campaign Analysis

It might seem incredible that, with spam the bane of our collective existence today, email marketing still seems worth pursuing. Not only is it worth it, but done right, email marketing can be one of the most productive acquisition channels for any company.

Analyzing email campaigns requires three important insights:

- You must use metrics that are unique to the medium.
- You can't track everything.
- You need to think end-to-end and not just in a silo.

As with paid search analysis, first you realize that you must deal with multiple sources of data. Your first source will be your campaign data, which in this case is your email provider. Your second source of data is your website.

Ensure that you have a tagging or tracking strategy for your email campaigns that allows you to merge the data after the fact. When I say your website data, I mean your broader data in Omniture, Webtrends, or Yahoo! Web Analytics. Tagging your site with the JavaScript tag from your email provider won't accomplish that. You'll have to ensure your campaign tracking parameters are set up so that the data will be captured by your web analytics tool.

You can break the analysis down into three important pieces: campaign response, website behavior, and business outcomes.

Campaign Response

The campaign response step involves the initial part of the customer experience: an email blast from you to get a customer response. The key metrics that you'll analyze will come typically from your email service provider (make sure you check that you can get these metrics before you sign up and fork over cash). Here are the metrics, beyond the standard emails sent, that will help you analyze performance:

Delivery rate = (# of emails sent – # of bounce backs) / # of emails sent This is your bread-and-butter Outcome metric when it comes to your campaign; it answers the following simple question: did we stand a chance at success? Note that the increasing use of junk and spam boxes means that bounce backs are not the cleanest way to measure deliverability. The emails might have just ended up in the junk email box where they never stood a chance of being opened.

Open rate = # of emails opened / # of emails sent Be aware that this metric is usually just directionally accurate. Most email programs now have preview panes that typically block images and scripts. Because of concerns about viruses and the like, the default settings on most email programs, including web-based ones, is to block images.

Click-to-open rate (CTOR) = # of clicks / # of emails opened

This is a key measure of the quality of your email list and of the effectiveness and relevance of your message. Segmenting this metric is really powerful. You can learn whether text messages or messages with images get a higher CTOR. You can compare customers in California, Idaho, and Florida; new and existing customers; or various demographics.

Subscriber retention rate = # subscribers – bounce backs – unsubscribes / # subscribers

This is the proof in the pudding, baby! This is perhaps as *strategic* an analysis as you could do for your email campaigns. Here you are measuring both the technical effectiveness of your email campaigns over time (reducing bounce backs) and the relevance of your messages and the targeting of the same (reducing unsubscribes). Measure retention

rate over time in aggregate—or for optimal health, segment retention rate—and measure it for the various objectives you have set for your email marketing program.

Website Behavior

In my humble experience, email campaigns usually measure the emails sent and Outcomes (Conversions), but they rarely pause and measure what happens once the person comes to the website. In part, this is because marketers are incentivized based on Outcomes, which is not a bad thing. But everything that happens on the site will either deliver high conversions or kill the most valuable offer you have ever sent.

Here are two useful metrics to inspire the kinds of analysis you should do for website behavior:

Bounce Rate = # of email campaign visits with a single Page View / # of email campaign visits

“Never let your campaigns write checks that your website cannot cash.” That’s really what you are measuring. Are your landing pages delivering on the promise you made in the email campaign? It does not matter if you have a 100 percent response rate on your email campaign if the website Bounce Rate is 99 percent. This metric helps you find opportunities for immediate improvement—such as pages you should test and calls to action and content that fail to deliver.

Length of Visit = percent of email campaign visits that last longer than xx seconds

Here’s the hypothesis: you are a content-only website, and you want people to come to your site and spend at least two minutes on it before they leave. You want to set a benchmark for the behavior of people who come from email campaigns. The xx in the definition forces you to think up front and plan for site behavior before you send an email blast and, of course, measure your performance against that goal. For a recent ecommerce client, I set a goal of three minutes and measured what percent of Visits from email campaigns exceeded three minutes. Why? For an average person, it would take six minutes from the time they landed on the site to learn about the product, compare options, start the purchase, and complete it. What the client wanted to know was how many Visits were *in the game*.

Depending on your type of website, other metrics will also be important to you. Do not ignore web behavior; it is an important part of measuring your email campaign success.

Business Outcomes

The next chapter is all about Outcomes and measuring success. That’s super important. Of course, you need to consider specific outcomes that affect your business when you run an email campaign. Here are some key metrics that you’ll analyze:

Conversion Rate = # of Orders / # of email campaign Visits

You need to measure Conversion Rate and segment it like crazy. Be ruthless at identifying causes for low performance. If you have a non-ecommerce website, no worries; you'll measure the # of *actions*, which could be lead submissions, # of Visits with more than nine Page Views, # of downloads, # of RSS sign-ups, and so on.

Average Revenue per Email Sent = total revenue / # of emails sent

I am very fond of this metric because it stresses productivity. Notice that you can play some really nice games by substituting the denominator with # of emails read or # of emails delivered. But I say, why settle for lower standards? Isn't the point of our email campaign to get maximum value? So, let's set a high bar; using # of emails sent will force quality in your email campaigns because it will mandate that the list be very clean and targeted. It is important in your journey as an analyst or marketer that you influence the positive behavior of your company by choosing your metric definitions wisely.

Email Campaign Profitability = (Revenue generated – campaign cost – cost of goods sold) / # of emails sent

Most email marketers will measure Revenue and Order Size and other such obvious metrics. But we rarely spend time measuring profitability. My analysis indicates that this is normally because it is hard to understand the true costs. In the case of email, that means the cost of the campaign: the cost of buying the list, sending the email, using resources, and so forth, as well as the cost of creating the products and services.

But that should not stop you from trying to measure Profitability along with Revenue. It is very easy to imagine that the most successful email campaign in the history of your company could very well cause bankruptcy (costs greater than revenue), and campaigns that look like potential losers could be most profitable. See why this is important?

Email marketing works. You just need to resist the temptation to abuse your customers. Don't preselect sign-up boxes and have an extra step to confirm opt-ins; always think of the customer benefit and not just what you will gain. You should have a big, clear, one-click unsubscribe link. Finally, you need to be relevant. That's really all it takes: treating your customers exactly as you would like to be treated.

Rich Experience Analysis: Flash, Video, and Widgets

Let's close this section with perhaps the toughest but coolest analysis you'll do: analyzing and understanding rich-media website experiences built with Ajax, Flash, Flex, or embedded pieces of rich content such as videos, widgets, and so on.

The challenge with analyzing these experiences is that they fundamentally *break* the core data collection mechanism in our web analytics tools, namely, the mighty Page View. Pretty much every analytics tool on the market today expects a Page View as a record of the interaction someone has with your website.

The problem? Most rich experiences don't generate Page Views. Consider something as simple as hitting Reply in Gmail: a box opens up at the bottom with no refresh

of the page. Or consider hitting the Play button on a video on your website. Neither of these actions generate Page Views.

The strategy of past analytics tools was to get you to configure these rich experiences to generate *fake page views*. That approach allowed the tool to fit data into existing structures and enabled reporting (conveniently without the tool having to retrofit itself for the new world).

Event Tracking

Today, some tools provide next-generation data collection models such as Event Tracking to capture data from rich experience. This data is collected differently from your standard tag-based Page View data, it is stored differently (no square pegs in round holes as with fake page views), and finally it creates new metrics that capture the unique experience of rich media.

Although the approach differs from the kind of web analytics tool you use, Event Tracking works by giving you a few *empty containers* for storing data (key-value pairs) according to a hierarchal model. For example, Google Analytics provides you with four empty containers, called Category, Action, Label, and Value. Let's solidify how you track rich experiences using a fictitious example; see Figure 4.45.



Figure 4.45 An example of a rich experience, on the Toyota website

The car configurator on the Toyota website is built using Adobe Flash; it is a beautiful *pageless* experience that allows Visitors to customize Toyota cars to their own preferences. The only way to do analytics on this experience is to step away from the traditional *page-driven* data collection mechanisms and use Event Tracking. The

only challenge with Event Tracking (or tracking rich media experiences) is that you have to spend some time up front defining success. This is something new for analysts and marketers, who are accustomed to just getting data spewed at them.

Ideally, Toyota wants to know which cars customers choose to customize, what the most common features people add to their cars, and what optional elements are popular. Figure 4.46 shows how you might create an optimal data model for collecting this type of data.

Object	Action	Label
CarModel	Transmission	Manual, Automatic
	Color	White, Flint, Silver, Black, Crimson, ...
	Exterior	Ground Effects, Fog Lights, Yakama Rack, ...
	Interior	Security, Sport Steering Wheel, ...
	Sound	Navigation, Pioneer Premium Audio, ...
	Performance	Quick Shifter, Lowering Springs, ...

Figure 4.46 Car configurator Event Tracking data model

Creating this data model up front allows the developers of the application to ensure the proper encoding is in place up front.

As you can imagine, the data model would be different based on the kind of rich experience you build. For example, if I have a website with lots of movies, then I want to track which movie my Visitors watch, where they pause, where they stop, or whether they watch completely. Since the movie is in pieces, I want to know which parts people watch and which they skip. Finally, given that a lot of the world has a slow Internet connection, I want to track how long it takes each part of the video to load. Figure 4.47 shows my potential Event Tracking data model.

Object	Action	Label	Value
Movie Title	Play	Part 1, Part 2, Part 3, Part 4 Part xx	Video Load Time
	Pause	Part 1, Part 2, Part 3, Part 4 Part xx	
	Stop	Part 1, Part 2, Part 3, Part 4 Part xx	
	25%	Part 1, Part 2, Part 3, Part 4 Part xx	
	50%	Part 1, Part 2, Part 3, Part 4 Part xx	
	75%	Part 1, Part 2, Part 3, Part 4 Part xx	
	100%	Part 1, Part 2, Part 3, Part 4 Part xx	

Figure 4.47 Video website Event Tracking data model

The amazing thing is that the developer only has to encode the video player on the website once, and with that, all movies played become trackable. There is no need to encode all 2,500 movies on the site (well, not unless you like pain!).

You report and analyze this data through standard reports available in your web analytics tool or by extracting the data using the API and putting your own spin on it. For our car configurator example, a standard report for the actions taken by website Visitors would look like Figure 4.48 (not Toyota's data).

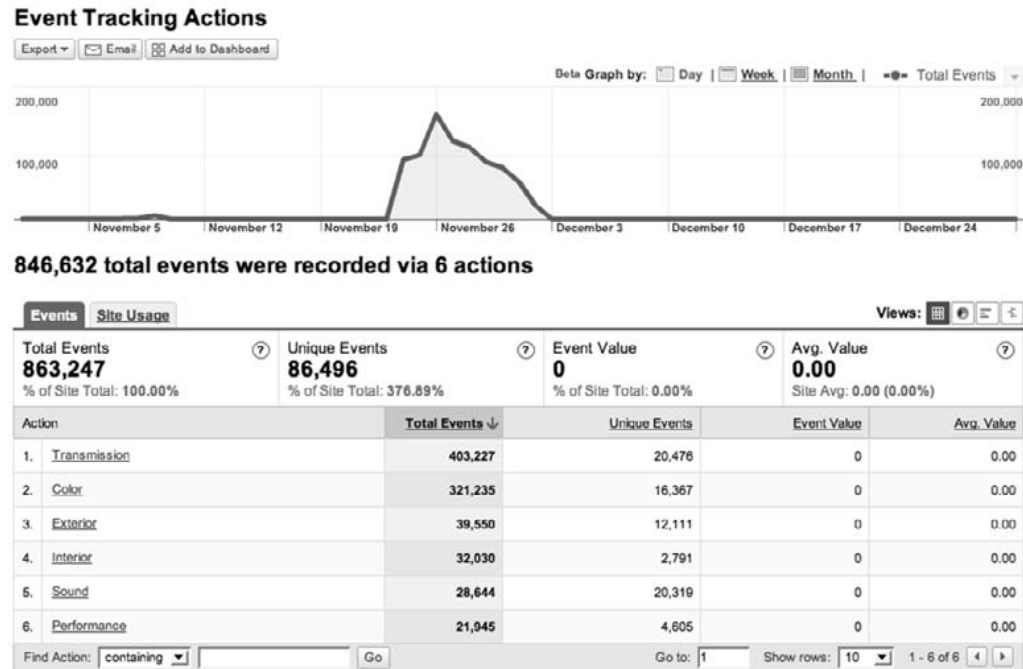


Figure 4.48 Options chosen by website Visitors

The configurator was live on the site for a few weeks. During that time, the most popular option customers chose was Transmission, followed by Color. The car company had expected Visitors to customize something as personal as the color of the car first. Hmm...it's time to rethink the kinds of customers we have!

In our movies example, standard metrics would report the number of videos played on the site, the number of visits with video plays, the average number of videos played per visitor, the number of videos completed by Visitors, the funnel of how many people start and reach the end, and how far they watched if they did abandon the video (25 percent, 50 percent, 75 percent, and so on).

That's just part of the fun. Then you can dive deeper into your data and segment it! You can see preferences by geography (what color of cars Californians prefer compared to Texans) or for current customers vs. new customers—the sky is the limit.

These new mechanisms to collect data are quite malleable. For example, you can encode widgets (that run around the Web getting all viral and whatnot) and track behaviors or new metrics that might not otherwise be possible. A good example is tracking how many widgets are *grabbed* (the action of taking a widget and placing it on a new website), thus taking a step toward computing *viralness*. Since widgets are essentially small, built-in browsable experiences, you can compute new metrics such as Interaction Time, the amount of time in seconds a Visitor on any site integrated with the widget (using mouseovers, clicks, and so on).

In summary, Page Views are dead; the future of the Web will be substantially richer and immersive. To track these experiences, we must think of new ways of collecting data, understand up front what success means, and ensure all encoding is correctly done at the start. By keeping an open mind about metrics, you'll use them to measure success, and I guarantee glory will be yours!

Reality Check: Perspectives on Key Web Analytics Challenges

You'll realize very quickly that some issues in web analytics will become the bane of your existence. They will hound you in your dreams. Your bosses will throw hissy fits about them. You will come to question why you started with web analytics in the first place!

This section of the book is my attempt at truly *ninjabfying* your mind. In other words, I want to get you a lot more focused, to appreciate value, and, perhaps most importantly, to selectively ignore distractions during your quest to find actionable insights.

Cookies, sampling, history, perfection, video playback, data reconciliation. Pay attention, grasshopper....

Visitor Tracking Cookies

Perhaps no other topic is more full of FUD (fear, uncertainty, and doubt) than cookies. Tracking cookies help fulfill a deep-seated desire in marketers and analysts on the Web: to track unique people who come to the website. But short of people embedding RFIDs in their bodies, which are scanned by every computer in the world, there is no way to track unique people.

So, as an ecosystem, we do the next best thing: we try to identify unique browsers (as in FireFox, Chrome, Internet Explorer, and so on) used to visit our websites. *Cookies* are small text files containing an anonymous unique identifier that stitches together Visits to the website by the same person (with the *assumption* that the same person uses the same browser).

We can do a lot of tracking without cookies; they are not the be-all and end-all of Visitor behavior tracking. But we currently track Unique Visitors using cookies. Some websites that make people log in to use key features, such as a bank, can use the

login ID to count Unique Visitors, though that ID tracks only those people who log in—and many people won't because they are not there to check their accounts.

Let's attack the rest of this complex issue in a few bite-sized, understandable chunks.

Transient vs. Persistent

The web analytics software will set two types of cookies when you visit a website. They are commonly called *transient* and *persistent* cookies. Some folks refer to them as *session* and *user* cookies, respectively.

The job of the transient cookies is to help *sessionize* an experience on a website. Put simply, your Visitor will make a series of clicks and leave. That's a session. The transient cookie helps group those clicks efficiently.

The transient cookie is *set* when a Visitor visits the site, and it disappears when he or she leaves. The persistent cookie is set the first time the Visitor visits the website, and it remains there for a duration determined by the website (18 months is common, though some tools will set cookies for 50 years). Persistent cookies help identify a unique browser to your website, inasmuch as they are the closest thing to tracking a person or unique visitor. The persistent cookie is on your browser until you either delete it, reinstall your browser, or perform similar actions.

Note: Persistent cookies don't contain any personally identifiable information (PII) data. They just have a random string of numbers or alphabets that only the company that set the cookie can read. For example, here is the identifier used to track me by a cookie that Webtrends just set on my browser as I visited www.webtrends.com: C8ctADY1LjU3LjI0NS4xMS00MTU3MTQwMTc2LjI5OTQ0NzE5AAAAAAAAAACAAaM0AAINghUgWYIVl.



First Party vs. Third Party

Transient and persistent refer to the nature of cookies. *First party* and *third party* refer to the type of cookies.

A *third-party* cookie is set by, well, a third party when someone visits your site. For example, when you visit www.omniture.com today, the tracking cookies are not set using the www.omniture.com domain; they are actually set from www.2o7.net domain. That makes the tracking methodology used by Omniture *third-party cookie* tracking.

In the good old days, web analytics could more easily use third-party cookies, and such cookies were rampant. But other players then used these cookies in suboptimal ways. This led to default Internet browser settings that rejected third-party cookies and many antispyware and malware programs that autodeleted them. Suffice it to say, third-party cookies have fallen out of favor, and they are considered suboptimal for tracking Unique Visitors.

A *first-party* cookie, hence, is set by the web analytics tool using the domain of the website itself. For example, when I visit www.coremetrics.com, I see that it set cookies using the domain data.coremetrics.com—which makes these cookies first party.

First-party cookies are the preferred tool of choice for tracking Unique Visitors because they are deleted or rejected a lot less by any objective measure. Therefore, they are far superior at tracking Repeat Visits or the New and Returning Visitor segments.

First-party cookies are also rejected a lot less because much of the Internet does not work if you don't accept first-party cookies. Email providers such as hotmail.com or gmail.com, ecommerce websites such as amazon.com or crutchfield.com, banks, or even blogging platforms all require you to accept first-party cookies.

Almost every decent web analytics vendor now provides an easy way for you to use first-party cookies. Some, such as Google Analytics, offer only first-party cookies. Others, such as Omniture and Webtrends, give you a choice.

If you notice some initial pushback from your vendor to use the *easier-for-them* third-party option, push back yourself. Insist on first party. It's good for your health.

Exception for Third-Party Cookies

There are some relevant uses of third-party cookies. One of the most common uses is by ad-serving platforms, because third-party cookies are the only way they can track a *unique browser* across multiple websites. So, even if that third-party cookie gets blown away and rejected more often, they (you) really don't have much of a choice. That's just how the Internet protocols work.

Here's an example of how third-party cookies work on an ad-serving platform:

I notice that omniture.com is using .2o7.net third-party cookies. After going to omniture.com, I could go to ebay.com and then to nytimes.com. The .2o7.net cookie knows that I was at the Omniture site a little while back and knows I then went to eBay and then NYTimes.

Now, as I am reading the latest Maureen Dowd column, the .2o7.net cookie could serve me an ad for Omniture next to the Maureen Dowd column. Knowing I also went to eBay, it could even give me a deal on Omniture in that ad!

This is of course just one example to illustrate the use of a third-party cookie and why Atlas and DoubleClick and Yahoo! and all the others use them (and provide value to their customers).

First-party cookies can't be "read" and "carried over" as in the previous scenario. Their anonymous data is restricted just to that one site.

Cookie Choice and Data Storage

Whether you use first-party or third-party cookies, it does not influence where your data is stored. The type of web analytics software you use determines that.

If you use an ASP-based solution (say, NetInsight, Yahoo! Web Analytics, or XiTi), then both your first-party or third-party cookie data is stored in the data center of your application service provider (vendor).

If you use an in-house solution (such as ClickTracks, Urchin, or Webtrends), then your data is stored in your own data center, regardless of the kind of cookie.

Cookie Deletion Rates

Remember that cookie rejection is not the same as deletion. With rejection, the visitor browser didn't even accept the cookie, which worsens your ability to track. With deletion, you get to collect data for the session (Visit), but you worsen tracking for subsequent visits.

Everyone wants to know cookie deletion rates ("Help! My web analytics data is stinky!"). Bad news: there is no global standard. I have never seen an objective study, that is, a study that was not pushing the vested interests of the publisher. It is also extremely difficult for a third party—an external company or agency—to gain the kind of access required to actual data that would help them develop anything close to an objective standard.

The biggest determining factors for cookie deletion are your customers, their browser settings, and the software on their computer. And those factors can vary greatly from site to site.

My experience of measuring cookie deletion using the company's own analytics data across a number of ecommerce, support, and other corporate sites has helped me come up with a "benchmark" of cookie deletion rates of 3 percent to 5 percent for first-party cookies and 20 percent to 25 percent for third-party cookies. They all tend to fall in that range.

But it is critical to realize that your number will be unique to you. To find your number, you need to put in the sweat, blood, and tears to measure it on your actual site.

The most common methodology for measuring cookie deletion involves extracting your JavaScript-tagged data from the analytics tool and doing exhaustive analysis of visit patterns in that data vs. your website server log files. Each set of data tracks something unique, allowing you to identify Unique Visitors. For example, your server logs have the `user_agent_id` as well as IP addresses. If you also allow people to log into your site, you can use that piece of key data in this analysis.

If you want to know the exact number of cookie deletions, you can't just take someone's word for it. You need to evaluate your own web analytics data and get your own benchmark.

Without Cookies, Life Continues

If you use cookies, then your key metrics such as Unique Visitors and Visits will be more accurate. Not perfect, just more accurate. You will also get a better

understanding of *pan-session* metrics such as Visits to Purchase or New and Returning Visitors or even Conversion Rates.

But if your company executives (the HiPPOs) or, more likely, website customers prefer that you not use cookies, then you don't have to use them. You won't be able to measure some of the previously discussed key performance indicators, but you can still get good value from the cookieless data that you do collect: Top Visited Pages, Revenue, Referring Websites (URLs), Search Engine Keywords, and so on.

Don't let the fact that you don't use cookies get in the way of using web analytics data in meaningful ways. Don't mope around! Use what you have, and you can still find actionable insights.

Data Sampling 411

Our minds usually refuse to accept that taking less data might sometimes be a better option. Hence, we tend to have an almost allergic reaction to hearing, "This is based on sampled data." We rarely accept that a slightly imperfect answer in 15 seconds is better than the perfect answer in 2 days. It is important to keep an open mind on this topic and realize that there are nuances to sampling—some very good and some you should use with care.

Pay special attention to the following two important challenges that drive data sampling:

- Almost all tag-based (and some log-based) paid web analytics tools bill you based on *pay per page view*. So, the more successful you are, the more you pay your analytics vendor (or often pay overages on top of your agreed Page View limits).
- You collect a ton of data and millions of Page Views, and now your simplest reports are slower than a snail on a hot day. Or you run massively complex, segmented queries against long time periods. Same snail outcome.

No web analytics vendor admits that either problem afflicts them. They will never tell you that data sampling might be important.

The standard operating procedure in dealing with these two challenges is to do data sampling, either selectively collecting data or selectively processing data. Although this practice is common, little is understood about its implications. Until now!

There are three primary ways of sampling your data:

- *Code Red*: Sampling web pages on your site
- *Code Orange*: Sampling data collected from each page
- *Code Green*: Sampling data processed when you run the query or report

Let's look at each approach in some detail.

Code Red: Sampling Web Pages on Your Site

Under the Code Red option, you add the JavaScript tag to only some pages on your website, either by choice or on advice from your vendor. Typically you might add the JavaScript tags to a bunch of your busiest pages and forget the rest (or as your CEO would say, “We should at least track our important pages even if we can’t afford to track the site!”).

This web page sampling is the least palatable of your three options. If you ever want to know anything interesting about your pages, you’ll run into a dead end because you simply have no data!

You will not have a complete picture of your website. For example, if you forgot to tag page *x* and your marketing department sent off a million direct marketing emails pointing to that page, or if page *y* was indexed by Google and attracts a bunch of traffic, you have no idea.

Code Orange: Sampling Data Collected from Each Page

In the Code Orange scenario, rather than collecting every single Page View on your site, you can specify in the JavaScript tag code something like this: *just collect every 10th Page View*.

You do this to reduce the bill you get from your web analytics vendor. Data will be collected every 10th time the page loads, and that data will be sent to your vendor. This means less data is collected for the vendor to store and process, and when reports run, you get that sampled data.

Now, in the report you will have “lower” numbers than your real numbers, but usually some approximation is applied (say multiply every number by 10) to get the “correct” numbers for you.

This is a suboptimal approach for so many reasons, but it is still better than collecting no data at all. In this case, you have some representative data for all your pages. Even with the multiplier, you get an approximate view of your overall metrics.

For pages that don’t get lots of Page Views, such as the pages beyond your top 20 or so pages, it also means that if you segment data, the quality of your reports will deteriorate very quickly. It is also suboptimal to sample Page Views; it is consistently better to sample sessions (Visits).

If you have a choice between Code Red and Code Orange, always choose Code Orange. Though you really want green!

Code Green: Sampling Data Processed When You Run the Query/Report

All data from your website is collected and stored by your vendor. The core challenge you face is getting fast results for your reports and queries, especially those that cover longer time periods or that perform complex queries.

There are two scenarios here. First, your web analytics vendor allows you to select the amount of sampling you want to apply to your data, which is a case of choosing your own *poison*, as shown in Figure 4.49.

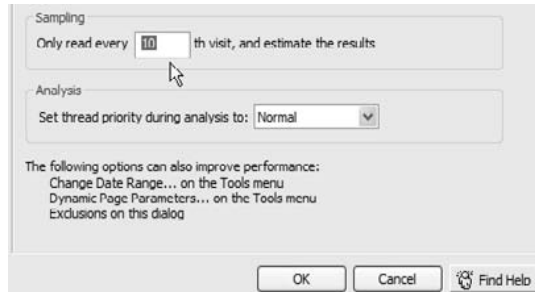


Figure 4.49 User-selectable sampling in ClickTracks

In this case, I tell ClickTracks to count behavior in every 10th Visit to my website (though I have all the data collected) to speed up the processing of the query. I get the answer I want much faster. If the count of Unique Visitors to my site last month was 18 million, then I might choose to sample every 999th session because, in the worst case, it will introduce an error of a few percent in my metric while providing me with the answer 500 times faster than if I crunched the whole thing.

In the second scenario, vendors such as Google Analytics don't have a user-defined sampling setting; rather, the tool will automatically trigger sampling when it detects that sampling will provide you with an answer quickly. Intelligent sampling is applied on the data, and your report might look like Figure 4.50.

At the very top of the report a message indicates that the report is based on sampled data. The table at the bottom shows the confidence intervals in the Visits column, which lets you know the *flex* in the data. For example, the number of Visits for New Visitors, 352,234, could be off by plus or minus 2 percent.

You collect all the data from your website, but rather than waiting hours to get your data back, both ClickTracks and Google Analytics allow you to get a fast answer within acceptable limits.

The ideal scenario is the one that ClickTracks uses, where I as a user can choose to apply the level of sampling on the data. For example, if I query hundreds of thousands of Visits over the past year, I might sample at, say, every 1,000th Visit. But if I query a very small segment of data, say everyone who came to my site on the keyword *avinash rocks*, I might set sampling at every ninth Visit. And if I really want to wait four days for my answer, I can run the raw unsampled query.

Most web analytics vendors are opaque about the kind of sampling they use (or impose on you) when they are trying to sell you their tool. Your job is to ask pointed questions and understand what type of sampling is applied, when, and how much control you have over it.



Figure 4.50 Statistically sampled report in Google Analytics

If your vendor gives you a choice, always choose Code Green over Code Orange and Code Orange over Code Red. But try to avoid Code Red; remember, in that case, you are guessing up front which data you don't need at all—and that's always unwise.

The Value of Historical Data

We have all been brought up to cherish data—to love it, marry it, and stick with it for better or for worse, to build increasingly vast and complex systems to keep it around, and to tap into cloud computing along with a small army of people in your company to keep that data happy.

This reverence sounds like a good mind-set, especially in the traditional world of ERP and CRM systems. But on the Web, this reverence can be a deeply suboptimal mind-set.

You need to divest yourself of the mind-set of keeping web analytics data around forever. If you, or your HiPPOs, have this mind-set, then a quickie divorce might be helpful.

Data Decay

Keep in mind that from the moment you collect your web analytics data, it starts to decay and lose value. Oh, it is useful on the first day and for the first month, but it's less so in six months. And click-level data is nearly worthless in a year.

Surely, you want an explanation! Sure. Let me give you six:

Your Visitors change too much Almost all the data you collect is anonymous, nonpersonal data from your Visitors. These darling Visitors swap browsers and machines and upgrades both from time to time. Thus, the data is less useful in identifying any usage trends and patterns tied to a unique person the further back in time you go.

Yet you make an implicit decision by hoarding data forever, hoping that you can use the data to make intelligent decisions about someone. Sadly, not gonna happen.

Your computations change too much You shift from logs to tags, and the core analytics data such as Visits, Visitors, or even Total Page Views now uses new and different logic. Even if you shift from Coremetrics to the capable and free Yahoo! Web Analytics, so from one tag-based solution to another tag-based solution, your data is not comparable because each tool uses its own logic to compute basic metrics or KPIs such as Conversion Rates.

Vendors and practitioners change their basic formulas for measuring core stats every so often. They rarely reprocess history (too hard!), making it difficult to provide continuity. As you evolve, comparisons become more like comparing apples and monkeys. There may be a relationship, but it's not one you're interested in knowing.

Your systems change too much At the end of the day, three pieces of information are captured by your analytics tool: the referrer, the page URL, and the cookie.

As you evolve your website CMS platform, say from Interwoven to ATG, or as you move it around or add or remove functionality, such as internal search or behavior targeting, you usually impact all three of those critical pieces of information that make up your data.

Resulting impact can make your data disjointed. And I am not even touching on changes such as moving from static HMTL to dynamic HTML or from personalized content to rich Internet applications (RIAs). All of these changes again impact the three pieces you collect.

Your website changes too much Google's home page has not changed very much in the last 10 years. That is the only exception I can find to the rule. However, your home page from three months ago is not your home page now (is it? I hope not!). You've killed half of your product-line pages last year, opting for product-detail pages for SEO reasons. There was no PayPal last week. Maybe 2009 was your first year with the support and ecommerce sites merged into one.

In the past six months, you have learned more about your business, about your data, about your Visitors, and about where you stand against your competition. Your web

presence has changed accordingly. Every change changes the data you have, and it changes the value you can get from data three months from now.

Your people change too much Sad as it may be, the hardest people to find now are web people—not just great web analysts, which we know are scarce, but web people in general. Front end, back end, thin, plump, newbies, experienced...all kinds are hard to find.

As people come and go, their actions have a subtle but important impact on all aspects of your data ecosystem. Given all these factors, would you still keep terabytes of data from two years ago? Should you invest 18 months building a data warehouse for your web analytics data?

I strongly encourage you to think of the pace of change. The chief reason for why historical data on the Web is not all it's cracked up to be is the decay.

The Goodness of Not Worrying About History

If you don't worry about history, then you are not tied to the past. You can think smart and move fast. If your current data will have less value in the future, then you will cherish the now more and try to get something valuable out of it.

Letting go of history also gives you the freedom to sever ties to legacy systems, legacy tools, or legacy data. You can move forward to better systems, tools, and data much faster than our sisters and brothers in the traditional world.

You can have a lot more fun because you get to learn, adapt, get value, and move on. It is damn exciting and damn liberating!

Saving What Matters

You do want to keep some history around, such as Weekly Trend (counts) of Visits and Unique Visitors, Top 10 Referrers to your website by month, and monthly Bounce Rate. You want to hold on to aggregated data for your critical few metrics, because that won't become less important with time. And keep some revenue stats just to prove that you're worth it!

Keep that data around as long as you have it. It will all fit on one tab in an Excel spreadsheet. That is all you'll need. Your historical data choices might be slightly different than the ones I mentioned, but I assure you it will fit in a spreadsheet.

To summarize, you should keep your click-level, detailed data around for a year (assuming seasonality!) and your session-level aggregated data for as long as you want or must, but it should fit in a spreadsheet.

It is extremely difficult to get anything out of your web analytics that you can take action on right now. I humbly recommend that in the drive to conquer history you don't forget the present. Don't ignore the price that you'll pay every day in the future for every day that has gone by.

History is important in other contexts, but in web analytics tools, for now, change on the Web reduces value from old data. This might cease to be the case at some point, but until that faraway day, you know what to do!

The Usefulness of Video Playback of Customer Experience

Video playback seems to make so much sense on paper: you can watch every Visit to your website like a sitcom. Sit back, and see what that person is doing! Finally, you can make more intelligent decisions.

The reality is that, except for one narrow and expensive case, video playback rarely produced the promised ROI. I especially wanted to cover this particular web analytics phenomenon because it illustrates the challenge of translating paper promises by vendors, or even our own marketers, into actionable insights.

The promise from companies like Tealeaf, ClickTale, RobotReplay, and others is simple: they will record all sessions on your website. You can then watch them at your leisure and figure out what's wrong with your website (or ideally what's great).

The following are a few challenges you face in finding insights.

Scale

If your website gets 100 Visits, then it is easy to do *random dips* into the data (videos), get a sense for things, and even get lucky and run into a problematic session. But, how do you do that if your website gets 100,000 Visits a month? How about 1,000,000?

It is incredibly hard to *get lucky*. Yes, you can still say this: "Let's find all sessions that end in an error and now narrow things down to those 1,800 sessions." But then, how do you watch all those sessions? How many hours are in the day? How many people do you need to watch the sessions to find something of value?

The promise of *individual* tracking is precisely what works against these solutions. It's a challenge of finding the signal in all the noise.

Interpretation

Video playback tools are sold as tools that perform usability. After all, you are watching someone doing something on your website. The problem, however, is not the watching; the problem is you: you and your interpretation of what you are watching.

When you study usability (remote or onsite), you have the context of what the person is trying to do in their own words. You are clear about their frustrations or about why they are failing because they tell you what they are trying to get done.

How do you do that in a *contextless* video? You guess. You impose your own interpretation on the data. Is it a surprise your insights are less than optimal?

If you notice that someone clicked Add to Cart and in the video you see a 404 error page, you don't need to interpret much: that's a problem. But you don't need to watch the videos of sessions to get that data. You could get the same data from Omniture or XiTi in two seconds by looking at the Next Page report from the cart.

“You Can Track Sara, Akio, and Roxana”

One of the most attractive on-paper values of session video playback is that if people call you or write to complain, then you can diagnose their experience. You can see what the problem on the site was.

But, here’s the underappreciated problem: how do you find Sara’s session on your website? How about Akio’s?

Most people on your website will have anonymous sessions, even as they enter secure pages like shopping carts and checkouts because they haven’t created an account. If Sara calls you on the phone, will you ask her to go into her browser, look for the cookies set, find the persistent cookie ID, and read off the 64- or 128-encrypted value that might possibly allow you to watch Sara’s session and diagnose the issue?

This is not a problem on a bank’s website. Visitors are forced to log in. When they call, a representative will ask for an account number or ID to find the problem session. Then the rep can watch the session and try to fix the problem.

Is your site a bank, or do you have a similar customer experience? Can you track “people” as they do? That answer is key.

If you are a small retailer, you can possibly “watch” all the sessions of people who added something to the cart and never purchased. Now you want to send those customers enticements to come back and buy. But then of course you’d need to find their contact information from the video of an anonymous session.

Cost vs. Benefit

A typical in-house hosted video playback service will require you to first buy the software from the company for a starting price of a couple hundred thousand dollars. Then you have to buy hardware in-house to host the software and service at a cost of another couple hundred thousand dollars.

You may go through months of implementation and integration processes with your company’s websites and systems. You will have to hire a part-time or full-time person to baby-sit this software and hardware implementation. You’ll hire more full-time analysts to make sure reporting works and to analyze some of the data (did you think I would forget that?).

That’s what it costs to get going. Now, what must you get out of this to justify the return on investment? How long will it take? Please give very careful consideration to the portfolio of decisions you have to make across your entire web analytics ecosystem (see Multiplicity in Chapter 1) and then prioritize your choice of video playback.

Of course, if you use something like ClickTale, RobotReplay, or others, you can reduce both cost and *time to data* because it is an ASP-based solution, hosted in the cloud rather than in-house in your company. In that case, you worry only about the first three challenges in this section.

Hope for the Future

At the moment, Video Replay's key flaw, other than cost and *time to insights*, is its over-reliance on people to find the signal in the noise and to interpret the data. Likewise, the anonymity of most sessions makes it really hard to find any signal in the noise.

My hope is that this will change in the future, and I have a simple idea of how. Why do I need to watch 200 sessions out of 177,825 to find out that there was a problem? The technology should find logical patterns in the videos and data automatically so that I can watch one *aggregated* session to find interesting customer behavior. ClickTale provides you with one video for all Visits in which a page existed, and the video shows an aggregated view of how far all people scrolled down a page. One video = aggregated data and value. We need a ton more of that from video playback vendors.

Here's another idea. Today in these tools, if any modicum of segmentation or rules-based analysis exists, it then falls on my favorite analyst to use her intelligence and considerable expertise to go through the massive mound of data to find insights. These systems should be intelligent enough to algorithmically mine the data to highlight three things I need to look at or do for my site.

At some point, the video playback solutions will become smart enough to do more than just "puke out data." They will apply technology and match raw computing power to ensure that people have access to key processed starting points. Then we'll start to get somewhere. Until then understand the challenges outlined here and proceed with caution.

The Ultimate Data Reconciliation Checklist

I believe in having one Clickstream web analytics tool and sticking with it, because it is hard enough to understand just one tool well enough to start finding insights and creating a data-driven organization. You throw two or three into the mix, and it becomes akin to Sisyphus pushing a stone up a mountain.

But the sad reality is many people have more than one tool on their website. We often practice bigamy when we should be practicing monogamy. That leads to the bane of every analyst's existence: data reconciliation!

In this section, I'll share a helpful and effective checklist of how to reconcile the data, but more than that, I'll share the nuances and complexities of online measurement. My fondest wish is that you'll appreciate the challenge better and be even more ninjalike in your thinking as you tackle other problems.

Figure 4.51 illustrates the challenge very nicely; you see data for the same time period from three different web analytics tools.

Notice something delightful—the number of Visits reported for the same time period varies across tools: StatCounter: 53,123, ClickTracks: 56,184, and Google Analytics: 45,710. It is a difficult and thankless task to reconcile this data, and rarely is it worth the effort, but when you have to do it, use the following effective checklist.

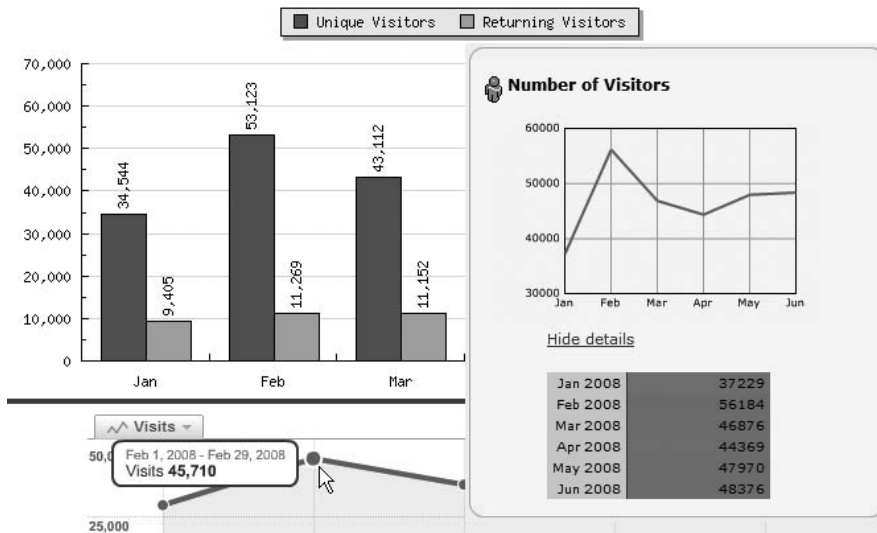


Figure 4.51 Visits measured by StatCounter, Google Analytics, and ClickTracks

#1: Don't Compare Data Collected from Web Logs vs. JavaScript Tags

Comparing web logs to JavaScript tags is like comparing apples and watermelons. It's just not right. These two methods of collecting data are very different, the processing and storage is different, and the things that impact are very different. Often, your numbers might not even come close.

Web logs primarily deal with effective and extensive filtering of robots (if you are not doing this, you are screwed regardless); the definition of a Unique Visitor (are you using cookies, just IP, or IP + user agent IDs?); and, this is increasingly minor, data caching at a browser or server level, which can also mean missing data from logs.

There is also the critical matter of rich-media content: Flash, video, Flex, apps, and so on. Without extensive custom coding, your web logs are clueless about all your rich-media experience. Most tag-based solutions now come with easy-to-implement solutions that will track rich media. So if you have a rich-media site, it will cause a big disparity between numbers you get from logs and numbers you get from tags.

The primary obstacles for JavaScript tags are browsers that have JavaScript turned off. Typically that will be 2–3 percent of browsers. In that case, the data will be missing from tag-based files but will exist in your log files. Be careful when comparing these two sources.

#2: The Gift That Keeps Giving: First- and Third-Party Cookies

I have covered cookie types a few times already, but the issue bears repeating. Using first-party or third-party cookies can have a huge impact on your key metrics such as Unique Visitors, Returning Visits, and so on.

Typically, if you use third party, then your numbers will be higher (and of course wrong), compared to numbers from your first-party cookie-based tool.

Cookie *flushing* (clearing cookies upon closing browser or by your friendly antispyware tool) affects both cookie types in the same way. Cookie rejection is more complex. Many new browsers don't even accept third-party cookies, which will impact your metrics.

Check your type of cookies; it will explain lots of your data differences.

#3: Imprecise Website Tagging

Other than cookies, imprecise tagging is the next most common source of data issues, and it shows up when you compare multiple JavaScript-based analytics tools.

The problem manifests itself in two ways:

- Incorrectly implemented tags
- Incompletely implemented tags

The standard JavaScript tags are pretty easy to implement. Just copy and paste, and happy birthday! But then you can add or adjust them to do more, which is why they can take eight months to implement. You can pass *sprops* and *evars* and *user_defined_values* and *variables* and *bacteria* (OK, kidding on that one).

You should make sure your Webtrends, NetInsight, or other tool is implemented correctly. That is, make sure it is passing data back to the vendor as you expect.

To check that I have implemented tags right and that the *sprops* are not passing *evars* and that *user_defined_values* are not sleeping with the *vars*, I use tools such as IEWatch Professional, Firebug, Web Developer Toolkit, and Web Bug.

Now, incompletely implemented tags are a simpler matter. Your IT department (or brother) implemented Omniture tags on some pages and Google Analytics on most pages. See the problem? Unfortunately, this problem is astonishingly common. Make sure you implement all tools on all the same pages, if not all pages on the site. To check that my tags are completely implemented, I use tools such as REL Software's Web Link Validator and WASP.

If you want faster reconciliation between your tools, make sure you have implemented all your analytics tools correctly and completely.

#4: Torture Your Vendor: Check Definitions of Key Metrics

In Figure 4.51 the numbers are different, of course, but you'll notice that each vendor has given its own sweet name to the metric: Visits, Visitors, and Unique Visitors. How exasperating!

As an industry, web analytics has grown organically. Each vendor has either created its own metrics or taken standard metrics and, just to mess with us, decided to call them something else. This naming inconsistency makes reconciliation harder.

In the following list, honest to God, are three definitions of Conversion Rate I have gotten from web analytics vendors:

Conversion: Order/Unique Visitors

Conversion: Orders/Visits

Conversion: Items Ordered/Clicks

What? Items Ordered/Clicks? Kill me now!

So, before you tar and feather a particular web analytics tool, take some time to ask your vendors what the precise definition is of the metric you are comparing. Often that can help explain the difference.

With some vendors, getting a definition can be hard. Vendors are secretive, protective, or even embarrassed. You might have to badger them. Do it.

#5: A Tough Nut: Sessionization

Sessionization is the process of taking a series of hits or requests to your website and identifying a Visit by the same browser. The problem is that different vendors do this in different ways.

Here is an example: I go directly to www.recovery.gov and see some pages, then I go to Google and do a search, and finally I return to www.recovery.gov. This all happens in a period of a few minutes.

In some analytics tools, this whole process will be considered one visit because it happened in less than 29 minutes. But other tools will see it as two visits because I went to a search engine and came back to the site. Some vendors will time out sessions after 29 minutes of inactivity, and others will do that after 15 minutes. Arrrrgh! See how session inconsistency makes reconciliation harder?

Another common issue arises when some tools set the “max session timeout” with a hard limit of 30 minutes while others set the limit at 12 hours or not at all. So, if you have long visits to your site, you may see one total visit or one visit broken up as ten.

Probe this important process because it affects the most foundational of all metrics, Visits.

#6: The Permanent Tripwire: URL Parameter Configuration

Life was so sweet when all sites were static. URLs were simple:

www.bestbuy.com/video/hot_hot_hottie_hot.html

Any web analytics tool could easily understand visits to that page and hence count Page Views. But then the Web became dynamic, and URLs now look like this:

www.bestbuy.com/site/olspage.jsp?id=abcat0800000&type=category

That’s the page on Best Buy for the Phone category, and here’s a page for a particular phone:

www.bestbuy.com/site//olspage.jsp?id=1205537515180&skuId=8793861&type=product

See the additional parameters? See everything after the question mark? As a final example, the following is the URL when I am on the page of the same phone but I have clicked a tab to read more details about the phone:

```
www.bestbuy.com/site/olspage.jsp?skuId=8793861&productCategoryId=abcat0802001&type=product&tab=7&id=12055375151
```

The problem is that while web analytics tools have gotten better and can probably understand that first page (phone category page), the next two pages are not as straightforward. These pages contain *tracking parameters* or *system parameters*. Junk. So, now you must sit down with your beloved IT folks and spend time documenting all the junk in the URL, in other words, the skuId, productCategoryId, type, tab, and id.

Some of these parameters make a web page unique, such as skuId, productCategoryId, and tab. That is, their presence and values contained mean it's a unique page. So, skuId=8793861 represents one phone, and skuId=8824739 is another.

But there will be some parameters that don't mean anything. For example, it does not matter if type=product is in the URL or not. Some of these different pieces of information are worth ignoring, and others you ignore at your own peril.

Your web analytics tool has a hard time taking all these pieces and counting the Page Views correctly. Each analytics tool has its own way of identifying and updating the settings for recognizing unique Page Views and ignoring the junk. If you want the data to reconcile, you'll have to make sure you do the same configuration in all tools.

If you are tracking rich media experiences using Event Tracking (as we covered earlier in this chapter) in one tool and generating "fake page views" in the other because it does not support Event Tracking, then your numbers will never match up.

#7: The Problem of the Big: Campaign Parameter Configuration

If you run lots of campaigns (email, affiliates, paid search, display, mobile, and so on), then you absolutely must tag your campaigns correctly and then configure your web analytics tools correctly. You do this to ensure your campaigns are reported correctly, your referrers are reported correctly, and your revenue and conversions are attributed correctly. This challenge typically afflicts larger companies.

Here is a simple example. If you search for Omniture in Yahoo!, you end up here:

```
www.omniture.com/static/278?s_scid=680217600000000309&clicksource=standard&OVRAW=omniture&OVKEY=omniture&OVMT=standard&OVADID=4822371011&OVKWID=130976483511
```

If you search for Omniture on Google, you end up here:

```
www.omniture.com/static/278?s_kwid=omniture|2109240905&s_scid=omniture|2109240905
```

You'll note that Omniture has done a great job of tagging their campaigns. Let's say Omniture uses Webtrends and Google Analytics for web analytics. So, the analyst

at Omniture must *configure* both tools with all the campaign parameters, the hierarchies, and whatnot. That will ensure that when the analyst clicks the Paid Search tab in the tool, these campaigns will be reported correctly.

You'll have to repeat this for your affiliate, email, display ads, and all other campaigns you have. If you use two tools, you'll have to do it twice. And each tool might not accept this data in the same way. For Webtrends, you might have to place it in the URL stem; in Coremetrics, you might have to put it in the cookies; in Google Analytics, it might have to be a customized JavaScript tag.

Suffice it to say that setting up all these parameters is not a walk in the park.

#8: The Hidden Angel: Data Sampling

I covered data sampling earlier in this chapter. You can sample at the source (not collect data at all), or you can sample at *run time* (collect all data but sample sessions/Visits). You want to ensure that you are applying the same kind of data sampling in all your web analytics tools.

The tragedy is that you rarely get a choice because the sampling is often hard-wired into analytics tools. In that case, you'll have to just accept that all, none, or some of your data will never reconcile.

But you should investigate and make sure the vendors are transparent with you. Once you have the information from the vendor, make sure you adjust your data reconciliation expectations accordingly.

#9: Order of Tags

The order of your tags, being the last item in the checklist, is not the hugest of deals. But on heavily trafficked websites, or ones that are just heavy, tag order can also affect the differences in the data.

As your web page loads, the tags are the last thing to load. This is actually a very good thing; you should always have your tags just before the [/body] tags. If you have more than one tag, then they are executed in the order they are implemented.

Sometimes on fat pages some tags might not get executed. This happens because the user has already clicked onward. Or it happens because you have custom-hacked the bejesus out of the tag, and it is now an obese tag and does not let the other, lean tags load in the time available.

If you want that least amount of extra checking, switch the order of your tags and see whether it helps. Tag order might explain the last percent of difference you are dying to get.

See what I mean when I say stick to monogamy?

Do data reconciliation when you have to, and now you know how. But my overall recommendation is to accept that data will be off. If two tools are within 10 percent of each other, then just move on to doing the analysis.

This page intentionally left blank.