Brian Clifton

SUCCESSFUL ANALYTICS

GAIN BUSINESS INSIGHTS BY MANAGING GOOGLE ANALYTICS

FOREWORD BY AVINASH KAUSHIK

Brian Clifton

SUCCESSFUL ANALYTICS

GAIN BUSINESS INSIGHTS BY MANAGING GOOGLE ANALYTICS

Advanced Web Metrics Ltd

To my family

Contents of ebook 1

👉 *This is ebook 1, ISBN 978-1-910591-01-7, containing chapters 1–5.Chapters 6–10 are in ebook 2, ISBN 978-1-910591-02-4, available from your ebook retailer*.

Contents of ebook 2

☛ *Chapters 6–10 are in ebook 2, ISBN 978-1-910591-02-4, available from your ebook retailer*.

Foreword

If you have $100 to *invest in magnificent, glorious success from your analytics efforts, invest $10 in tools and implementation and invest $90 in big brains (people!)*.

I humbly postulated that ground truth as the 10/90 rule on May 19, 2006. With every passing year, I've come to believe in that rule more and more (and more and more). The reason is quite simple. Every facet of the business world is throwing off ever more data, and every facet of our personal existence (and insistence on sharing) is throwing off ever more data. Data, it turns out, is free; identifying specific actions business leaders can take based on rigorous analysis is not free.

This is why I'm so excited about Brian's book. It dispenses with the normal *omg, omg, look at how much data is there and is that not amazing, let us spend 18 months on implementation*, and gets to what it really takes to shift from data puking to recommending business actions based on data.

Here's one of my personal examples of the difference in emphasis, and what ultimately drives success. In every company, every leader wants a dashboard. "Get me a summary of the business performance. Decisions shall be made!" Analysts scurry around and an intense burst of data, manifested as tables and charts, is presented on a vanilla-scented piece of paper. Happiness? Job promotions?

Sadly, no.

It turns out that the higher you go up the chain of command, the more analytical skills go down, and the context required to make sense of the numbers on the dashboard is also dramatically reduced. Few decisions are made, and if there is a meeting to discuss this it devolves into a discussion of the data quality, missing data, colors in charts, and everything except making a business decision.

The answer? Words in English. More specifically: insights, actions, business impact.

Every dashboard in the world should include as few tables and charts as possible. It should include insights written in English (or your native language) by the analyst, followed by the recommended actions and—the most important critical must-have bit—the impact on the business if the actions are taken.

That vanilla-scented piece of paper will no longer drive one more awful discussion about the data itself; it will drive a discussion of which actions to take first. Hallelujah!!

It is incredible to realize that in the end, data by itself does nothing. It is just data. It is the $90 part—the big brains—that identifies insights, actions, and business impact that will push your company's profitability and customer delight to

new, incredible, heights.

Next time you receive a dashboard, look for the balance between tables, charts, and English text, and you'll know if it will add value or waste time.

That's my little appetizer for you as you dive into Brian's wonderful book.

The entire book is awesome. It is beautifully structured, and you should go from Chapter 1 to Chapter 10 on your *we will make the most of data* voyage. But if you wanted to be a little naughty and jump around, my favorites are Chapter 8 (you can read it anytime, and you can't work on the recommendations soon enough!) and Chapter 10 (every time you find a task daunting, find hope in the success of others in the case studies).

I wish you all the very best. Carpe diem!

Avinash Kaushik

Digital Marketing Evangelist: Google

Author: *Web Analytics 2.0*, *Web Analytics: An Hour a Day*

Preface

This is my fourth book on Google Analytics, but this one is different. Rather than making it a tool-specific practitioner's bible (as my *Advanced Web Metrics* series endeavored to be), I approached this book as I do my work: helping ambitious organizations make a success of their business by using data intelligently.

As I have come to realize over the years, success does not depend on tool expertise alone. The bigger issue is the organization. It needs to trust the data and have confidence in the process, structure, and people behind it—things not directly related to the tools being used. So I approached this book very much from the business point of view first, then worked backward toward the nontechnical aspects of the tool—Google Analytics. My intention is that senior managers, stakeholders, and practitioners all speak a shared language and set a common path to building a credible data-driven environment. I hope the method has worked.

As for all authors, my writing of this book was not a solo exercise. It required love, support, help, guidance, advice, friendship—even random and unrelated conversations. (You would be surprised at what can spark an idea connected with data!) The people I list here are those who have directly contributed to the book or to my thinking about applying successful analytics.

Sara Clifton's never-ending love, support, and guidance keep me on the right track and always help me to see the bigger picture of measurement, digital, and life in general.

Shelby Thayer has sanity-checked every word of my last three books, including this one. She is a great analyst, with a ton of experience at driving web measurement acceptance within a large organization, and her feedback and experiences have helped me significantly in writing a tightly focused book.

Brad Townsend is my valued technical editor. He is a smart (and modest) Googler who, as a software engineer, knows the technicalities and back end of Google Analytics like no other. David Vallejo, an expert Google Analytics implementer, developer, and all-round smart guy, helped me enormously with his technical problem-solving skills. Nothing can't be done with this guy at hand! Dave Evans expertly reviewed Chapter 7 ("Data Responsibilities") and provided insightful discussions about data privacy law. Dick Margulis is my trusted editor, who has now helped me write and structure three books and is my go-to man for navigating the tricky waters of the publishing world. His guidance and advice have been invaluable.

Avinash Kaushik has honored me (again) by writing the foreword to this book and setting the scene so enthusiastically and logically for the reader—in a way that only he can. I am lucky to count him as a friend and former colleague. He inspires me (and many others) with his advocacy and excitement for all things that

can be measured.

John Wedderburn, Tobias Johansson, and the team at Search Integration (where I work) have engaged in many “quality time” meetings and open-ended discussions that have broadened and deepened my knowledge.

And last but not least, the vibrant and smart GACP community pushes back the boundaries of what can be done with Google Analytics, and importantly, what can be simplified with it.

I hope I have remembered everyone.

Brian Clifton

January 2015

Introduction

> "We think we want information when really we want knowledge."
>
> —*Nate Silver, from* The Signal and the Noise

According to a recent survey of IT professionals,[1] "**55% of big data analytics projects are abandoned**." Most of the respondents said that the top two reasons the projects fail are that managers lack the right expertise in house to connect the dots around data to form appropriate insights, and that projects lack business context around data. Similarly, the "Online Measurement and Strategy Report 2013" from **Econsultancy**[2] **asked companies, "Do you have a company-wide strategy that ties data collection and analysis to business objectives?" Only 19% said yes**, a figure that had hardly changed during the previous five years.

I wrote this book for those managers struggling to make headway—to empower you to make informed decisions and overcome the obstacles.

My goal with this book is to get you to think in terms of *insights*—not Google Analytics data. An insight is knowledge that you can relate to. It's a story that puts you in the shoes of your visitors, so that you can understand their requirements when they come to and view your website, app, or other digital content.

A company's ability to satisfy the needs of a website visitor depends on two important factors:

- Visitor expectations, discerned from how they got to your content—what search engine, campaign ads, or social conversation drove their decision to seek you out
- User experience—how easy it was to use your content, to navigate around, find information, engage with you (contact you, purchase, subscribe, give feedback)

It is your organization's ability to manage, analyze, and improve these two factors that determines your digital success (or not). In this book I describe how insights are used to pull *all* of the relevant data points together to build a story of your visitors' journey and their experiences. With that knowledge you can improve these: as I show in Chapter 10, improvement can be dramatic performance gains in terms of your online visibility, revenue, or efficiency savings.

Yet Google Analytics doesn't provide insights by itself—no tool can. Producing insights requires an understanding of your business and its products, your value proposition, your website content, its engagement points and processes, and of course its marketing plan. Google Analytics provides the data (and lots of it) that enables you to assess these. However, people—not machines—build insights. This is the role of your analytics team. They sift through the noise to find the

useful data, translate it into information to explain what is happening, then build stories of useful knowledge for the organization—the insights.

This book is about showing you how to do that. This book is about knowing what to focus on, what you can expect in return, the talent you need to hire, the processes you need to put in place, the pitfalls to avoid, and how much investment is required in order to make it all happen.

This is a detailed book by necessity. Building an environment where you can trust your data, understand it, and make important decisions based on it requires a deep level of immersion, not an executive summary. However, my approach throughout this book is to focus on the insight gained for the business, not the minutiae.

This book is for you if you are a manager who needs an overview of the key principles of website measurement, the capabilities of Google Analytics, and how to grow and give direction to your organization when it comes to its digital strategy. Your ultimate interest is in insights and knowledge, not more data!

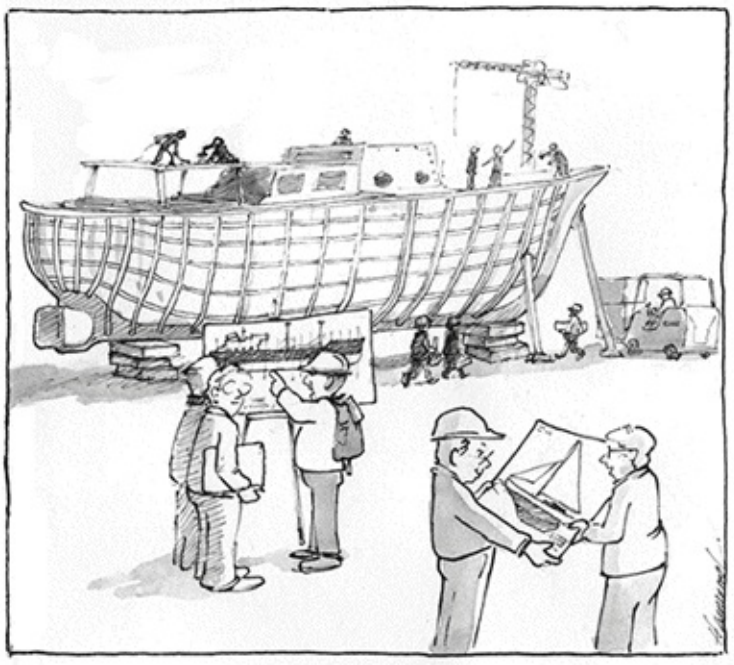
In short, I aim to put you in control and provide a perspective on the entire process of building a *data-driven* environment using Google Analytics.

REFERENCES

1 http://visual.ly/cios-big-data

2 https://econsultancy.com/reports/online-measurement-and-strategy-report

1 Preparing to Measure Success



You know measurement is important to your success—be it for you, your organization, or your career. In many respects, Google Analytics is just a tool, like the plethora of other data tools organizations use to help them make better decisions.

But web analytics—the area, technique, and industry that Google Analytics resides in—is different. Its reach and potential are far greater than any other tool you have. Why? Because not only can it measure the engagement, transactions, and revenue from your *customers*, it can also measure your *potential customers*—where they come from, what they are looking for, and how close they come to becoming a customer before they bail out (and where on your site or mobile app this happens).

This integration of customer and potential customer data is unique to web

analytics—and therefore extremely powerful. For example, the vast majority of websites have very low conversion rates—typically 3% (Figure 1.1).[1] That is, only 3 visitors out of 100 go on to become customers. While many a business analyst is tasked with optimizing for that thin sliver of a segment, there is clearly a much greater potential in understanding why the other 97% of people that expressed an interest in you (they visited your website) did not convert—and using this information to improve matters.

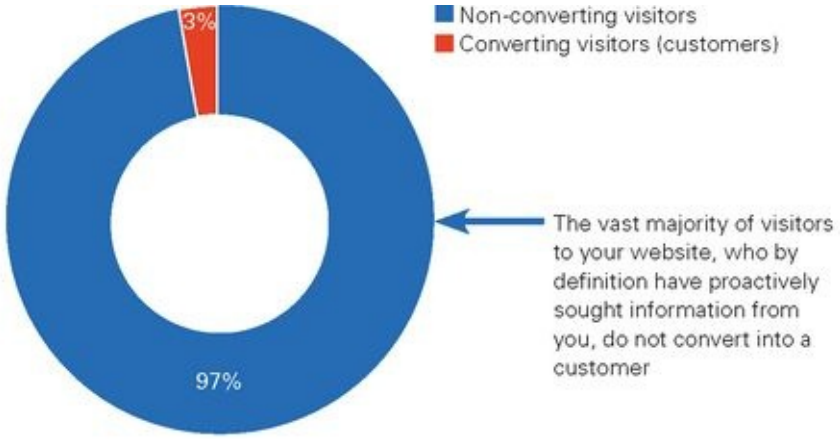


Figure 1.1 *The vast majority of website visitors do not convert, though most businesses focus on analyzing the small number who do.*

Google Analytics can be used to analyze both customer and non customer behavior. All that is required is a single digital touch point during their engagement with your organization. Usually, the touch point is a visit to your website, though with Universal Analytics (the latest enhancements to Google Analytics, described in more detail in Chapter 6, in ebook 2), it need not be. For example, a potential customer receives an offer from you via snail mail. This contains a coupon that they take into your brick-and-mortar store to make their purchase. At the point of sale, your store sends the purchase details (product name, value, coupon code, and so forth) via the Internet to your Google Analytics account. The result is that Google Analytics can generate a report on the performance of your direct mailings and sales in your store.

The digital touch point in this example was the actual purchase. If

a = number of direct mailings sent = 100,000

b = number of purchases with coupon = 725

then

campaign performance = *b* / *a* = 0.7%

Suppose your direct mail encourages recipients to visit your website *first* in order to obtain their coupon code. The second digital touch point is your website, and visits to it reflect the interest in your offer. Now you have a simple yet powerful set of data that can be analyzed—even if the recipient does not go on to purchase. For example, if

a = number of direct mailings sent = 100,000

b = number of purchases with coupon = 725
c = number of campaign visitors to your website = 8,000

then

interest level = *c* / *a* = 8.0%
campaign performance = *b* / *a* = 0.7%
website conversion rate = *b* / *c* = 9.1%

The extra data point collected in this second list (*c*) tells you that evaluating the results of a direct mailing is not as black and white as just the number of purchases: 8,000 people are actually interested in your offer; without this piece of information, it looks like only 725 people are interested. Armed with this extra data, you can now improve your direct mailing to increase that number—that is, grow the interest level. Simultaneously, you can work to improve your website landing pages to go beyond 9.1% conversion, and you can also try to increase the order value for those purchases (perhaps you can upsell and cross-sell related products). The result is that you now have a great deal more options to improve sales and measure the impact of your various efforts—so you can focus on the most profitable. Powerful numbers indeed.

Business Intelligence Defined

All analytical tools that help your organization understand itself come under the umbrella term *business intelligence*. Google Analytics is one of these tools. For clarity, I define three particular subcategories of business intelligence.

Customer Analytics The mining of existing customer data in order to discover buying patterns and demographic information. Often this information is used to compile a marketing campaign for the upsell and cross-sell of products to existing customers, as well as improve customer retention.

Web Analytics The study of your website visitor's online experience in order to improve it. The vast majority of data is completely anonymous. Google Analytics has traditionally been used as a web analytics tool.

Digital Analytics The evolution of web analytics to encompass all Internet-connected devices that can send a structured packet of data via HTTP, such as mobile apps, barcode scanners, checkout machines, stock taking, call center performance, or RFIDs. With its latest update, Google Analytics is now a digital analytics tool.

THE VALUE OF WEB ANALYTICS DATA

Consider the following exercise:

Many a business analyst is tasked with customer analytics (see sidebar, “Business Intelligence Defined”). Results can be used to compile a campaign for the upsell and cross-sell of products as well as to improve customer retention. Typical revenue increases are on the order of 1–9%.

Let’s invest the same amount of energy with a web analyst for the same organization. In this case, it is to understand the pain points of your website conversion process. The information gained is then used to reduce the friction of the process and improve the conversion rate—that is, generate more customers. In my experience, typical improvements are double digits, often triple digits (Figure 1.2).

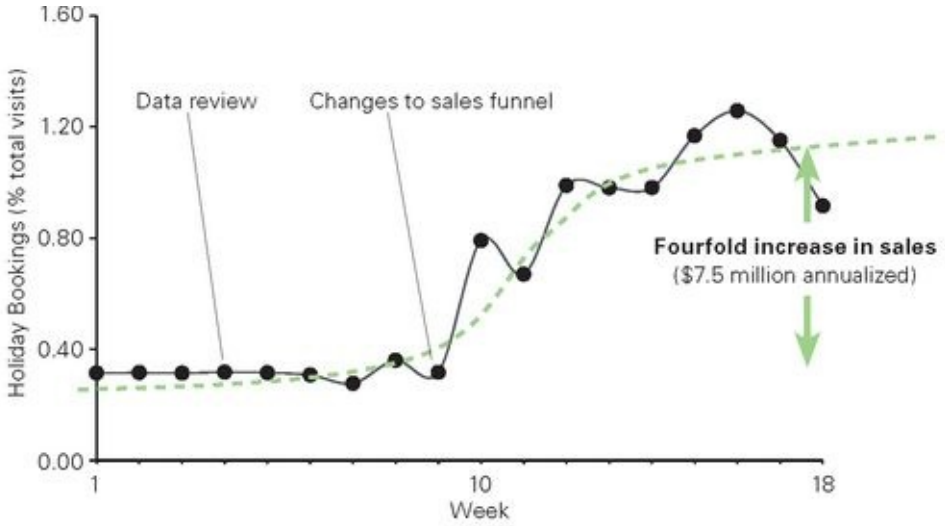


Figure 1.2 *Improvement in the rate of bookings for a travel company after studying their web analytics data*

In addition, the changes made to improve the conversion process on your website are compound. For example, if you make your website 10% more efficient at converting a visitor to a customer, that improvement is not just a one-off hit—it lasts perpetually (though not quite in practice). Therefore, in twelve months’ time when different visitors are coming to your site, 10% more of them will be converting than before. Additionally, a visitor who turns away in frustration because of a poor user experience is unlikely to return. But for *every* new visitor you acquire through an improved process, there is a better chance they will convert and become a long-term paying customer. The result of both of these is that your increased revenue will grow way beyond the initial uplift. Your marketing efforts just became a whole lot more efficient!

The huge potential of being able to convert more visitors into customers, or being smarter at acquiring higher-value visitors, is the great ability that web analytics brings to the table. And Google Analytics is a class-leading product in this field.

Suppose that by looking closely at how visitors interact with your website and

using techniques such as sales funnel analysis, exit points, bounce rates (single-page visits), and engagement metrics, you were able to improve your online conversion rate from 3% to 4% (a 33% improvement in the base rate). What would that mean for your bottom line? Let's look at a hypothetical example. If

v = number of visitors = 100,000

c = cost per visit = $1.00

then

a = cost of all visits = $v \times c$ = $100,000

The visitor acquisition cost is the same regardless of the conversion rate. The non-marketing profit margin, marketing costs, and revenue per conversion are also independent of the conversion rate:

m = non-marketing profit margin = 50%

s = marketing costs = $100,000

u = revenue per conversion = $75

Table 1.1 shows the results of improving the conversion rate from 3% to 4%.

Table 1.1 *The Economics of Improving Your Conversion Rate*

	Before	After
r = *conversion rate*	3%	4%
C = *conversions* = $r \times v$	3,000	4,000
T = *total revenue* = $u \times C$	$225,000	$300,000
n = *non-marketing costs* = $(1 - m) \times T$	$112,500	$150,000
P = *total profit* = $T - (n + s)$	$12,500	$50,000
R = *total ROI* = P / s	13%	50%

The last two rows of Table 1.1 put the analysis into context: profit will rise by $37,500 and return on investment will quadruple to 50%. Note that this is achieved solely by improving the conversion rate of the site—visitor acquisition costs remain the same. This is the value that web analytics can bring to your business. As a practitioner of over 15 years, I see this potential time and time again. It's real and attainable.

WHAT'S DIFFERENT ABOUT WEBSITE MEASUREMENT?

I often find that measuring the performance of a website is misunderstood by senior managers. That is not surprising considering that over the years the web analytics industry has struggled to define itself as it migrated from the IT department into marketing.

The principle of web analytics is straightforward:

> Web analytics is the study of online experience in order to improve it.

Describing Digital Analytics

Throughout this book I refer to *web analytics*, as this is the main use of Google Analytics today—analyzing website performance and its impact on other sales and marketing channels. However, that platform-specific definition that has existed since the 1990s is now beginning to erode. Users connect with your brand in multiple ways, be it through a traditional web browser, a mobile app, digital TV, or any other Internet-connected device capable of sending an HTTP request. An example of the latter is tracking the performance of your in-store checkout machines, or scanning badges as people walk into your event. In other words, there are lots of possibilities. The latest version of Google Analytics is capable of all these. More in Chapter 6 (in ebook 2).

However, managers who require data to guide their decisions are much more used to certainties—for example, the certainties that come from customer analytics about how much money you make, what the profit is, and how many customers you have, as well as operational analytics that tell you how many staff you have, what they cost, what your manufacturing costs are, and so forth.

By *certainties*, I am referring to hard numbers—solid numbers, where there is little or no error. If you wanted to know how much cash your company took in last month, you could simply print out your bank statement. That number is definite because it represents confirmed transactions. Your bank has done all the hard work to ensure that only valid payments are processed, transactions are legitimate, and the money sitting in your account is actually yours. Similarly, if you wanted to know the number of customers you have, you could make a query to your CRM system. That number is definite because it represents real people—the names and addresses of customers who have ordered and paid you. Your sales team has done all the hard work to ensure this is correct.

Things are very different when it comes to web analytics. This is because *all* your reported numbers are fuzzy, fluffy, hazy—in other words, inaccurate. Unlike the traditional business analyst, the web analyst needs to take responsibility for data quality.

Data Quality and Ownership

Data quality problems come from a variety of sources, the most common being an incomplete or poorly implemented setup: there is no bank or sales team verifying the data (your analytics team must do this, though I have found from experience this rarely happens). But even if you were able to attain a perfect setup, there are inherent inaccuracies. That's because the vast majority of collected data from your website is from anonymous visitors (the other 97%)—you have no idea who they are. Therefore, there isn't a one-to-one correlation of data to a specific person.

For example, if a visitor does not log in to your website or connect with you in some unique way, that visitor will be counted as a separate visitor to your website should they return using a different device—a tablet versus a smartphone versus a laptop. As far as your web analytics tool is concerned, this is counted as three different one-time visitors. The same effect happens if they use multiple browsers for subsequent visits, such as Internet Explorer, Chrome, or Firefox.

What about transactions?

You may feel that e-commerce data collected by your web analytics tool would fare better and be more accurate than anonymous visitor data. After all, a transaction is confirmed as completed, so you have the extra check taking place. That is true, but web analytics tools are poor at handling cancellations and returns because of what this actually correlates to. For example, a return of goods will cancel out the effectiveness of the campaigns that drove the visitor to your site in the first place. But if, say, I ordered the wrong size shoes from your website and returned them, the campaigns that drove my interest in your business are still valid, and should be credited for the purchase. So it makes more sense, from a marketing evaluation viewpoint, to keep *all* transactions within your web analytics data (except of course any fraudulent, test, or obvious error transactions).

How Accurate Is Web Analytics?

Don't worry, the inaccuracy of web measurement is something you get used to. It's an error bar. However, that error bar must be continuously monitored and corrected—it is not a set-and-forget operation. The web analytics team must take ownership of data quality. This is an important differentiator between web analytics and other forms of business analytics.

Assuming you have a good implementation of Google Analytics, your error bars should be within 5% of the true number. That is actually a very small error bar compared to the estimates that traditional marketers have to work with, such as newspaper circulation figures and TV viewing figures. See Chapter 4.

As with all data, accuracy is important. Even a perfect setup of Google Analytics will degrade over time as your website changes, the web changes, and user behavior changes. In order to trust your web analytics data and therefore have confidence in it to make important strategic decisions, you must take ownership of

data quality.

A final point on accuracy is the perception by many people, even smart people, that the vast volumes of data web analytics tools collect make the data accurate. My feeling is that most people are aware of the concept of small sample sizes yielding inaccurate results. Therefore, it is tempting to think that a huge pool of data drowns out any inaccuracies. However, web analytics data is so easy to collect that this common assumption is not true. The reverse can be the case: it is all too easy to collect inaccurate data and lots of noise, drowning out the important signals.

WHERE GOOGLE ANALYTICS FITS

Your website is in a unique position. As the first point of call for your digital presence, it is where your customers and noncustomers (potential customers, job seekers, investors, press, even competitors) go to find information on your business, products, and services. As such, web analytics is the only place where you get to see the data for *all* of these, side by side.

The side-by-side comparison of disparate data is invaluable for a senior manager. It lets you zoom out and see the bigger picture so you can use the same metrics and methodology to determine the relative success of each area—compare apples with apples, in other words. And that provides you with context—something often missing when analysts come to you with their deep-in-the-woods investigations. Context ensures you focus on the areas that have the biggest impact for your business.

That said, a single tool is not a panacea. Google Analytics (Figure 1.3) is in a unique position to connect your organization's digital activities with traditional offline marketing and your existing customers. Nonetheless, those individual areas still require their own tools for day-to-day management. For analyzing and understanding the performance of these other areas, use your central analytics platform—Google Analytics—as your unified measurement tool. Otherwise you will waste a great deal of time and energy chasing numbers from different tools that in principle should match but in reality never do. When it comes to counting, there are in fact differing techniques, methodologies, and definitions for the same thing!

Figure 1.3 *Google Analytics measures not only your customers' activity but also your potential customers' activity.*

Google Analytics collects and reports data. It is great for telling you *what* happened and *when*, but it does not tell you *why* it happened. That is where your analysts come in. A good web analyst uses her knowledge and experience to build a hypothesis to explain the *why*. Then she hunts down other data points to either support or refute the hypothesis— potentially outside of Google Analytics. If the data is inconclusive, a test (experiment) is performed. More on this in Chapter

8 (in ebook 2).

Why Google Analytics Is Not Customer Analytics

The dotted line between Google Analytics and customer analytics in Figure 1.3 is deliberate. I use it to emphasize two important points:

- The vast majority of web data (typically 97%) is anonymous, whereas customer analytics deals wholly with specific people and companies. Even assuming you have a separate login for your customers (therebyidentifying them), not all will use this when browsing your website. They remain anonymous even though they are your customers.
- When you track your customers with Google Analytics, the performance of your website is not the same as the performance of your sales. For example, your website may be great at converting visitors into customers. However, your sales may be poor because of returns.

You can do a great deal of customer analysis using Google Analytics. But bear these two points in mind to avoid overanalyzing the wrong area.

Data Is Not a Silo

Web analytics is different from other forms of business analytics—in terms of its potential (customer *and* potential customer data), data quality ownership (within the web analytics team), and the lack of certainty in the numbers produced (the large amount of anonymous data). But web analytics should not be treated differently as an information source within your organization from other kinds of analytics. Instead, I want to get you to think differently about how you treat all your data within your organization.

Most often data analysis is in a silo within an organization: it is set up as a separate department to be consulted with as needed. The people working in the data department are considered nerds by their colleagues—number crunchers, bean counters, analytical rather than creative (left brain rather than right brain) thinkers. As any data nerd will tell you—and I am one myself—collecting or studying data is boring. It's what you do with the resulting knowledge that is exciting and creative!

There is nothing more demoralizing for a team of smart people who have exciting insights to share than to find that none, or very few, of their suggestions or ideas are taken on board by the organization, that no change happens as a result of what has been learned from studying the data. This happens when the data is kept in its silo and the team is reduced to being report monkeys; it's a wasted opportunity for the business and a waste of resource talent.

By integrating your data, all data (though specifically I am referring to web analytics), within your marketing, PR, sales, web development, content creation, customer retention, and other teams, you will be able to create an environment for change. On the Internet, if you are not continuously evolving, even a market leader can be dead in the water within a couple of years.

Building an environment for change using a foundation of solid data is my approach throughout this book.

WHY GOOGLE ANALYTICS IS DIFFERENT

Two key things to be aware of about Google Analytics are its pricing model (free) and its dominant market share.

The Free Version

In November 2005 Google launched Google Analytics for free. That event created quite a stir in the web measurement industry (I was the head of web analytics for Google Europe at the time). Legitimate questions were raised: Why is it free? What's the catch? What will Google do with this client data? Can we trust the data provided by a vendor with a vested interest—that is, where we also advertise? (See "The Delicate Questions Answered," later in this chapter.)

What became abundantly clear was that before Google Analytics, price had been a limiting factor for measurement adoption. At Google we suspected this, but not to such a great extent—within one week of launch, more accounts were opened with Google Analytics than existed in the entire web measurement industry before the launch (Figure 1.4).

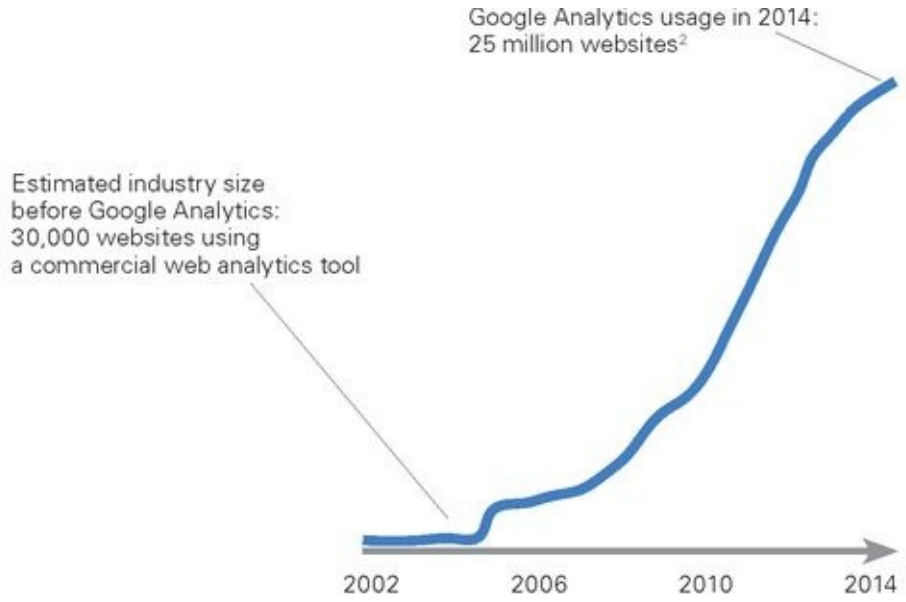


Figure 1.4 *The schematic rise of web analytics adoption since the launch of Google Analytics in November 2005*

According to builtwith.com, Google Analytics is now found on 25 million websites.[2] Free has clearly been an important driver in that adoption. So have its strong and continuously evolving feature set, ease of use, and large user community that openly shares knowledge and best practice. Those additional benefits have resulted in Google Analytics being adopted by many enterprises as well—firms where price is not a key factor in selecting the right tool for the job (Figure 1.5).

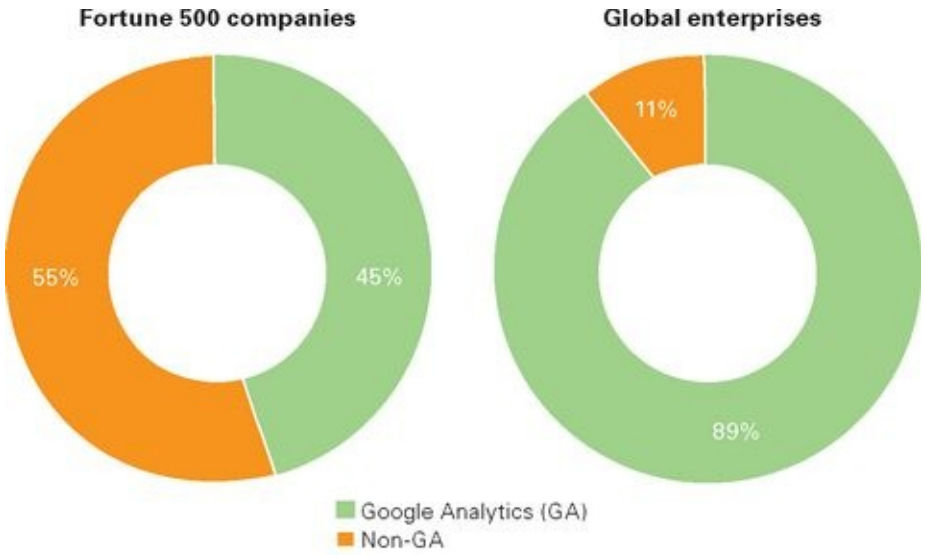


Figure 1.5 *The market share of Google Analytics*[3,4]

Google Analytics is a robust and user-friendly analytics platform that millions of businesses around the world rely on (it is available and supported in 40 languages). It is also an enterprise-class product trusted by brands that include P&G, Visa, General Electric, Sony, Toyota, Twitter, and the BBC. A longer list is available at http://brianclifton.com/who-uses-ga.

Google Analytics Market Share

There are two ways to estimate the level of adoption for Google Analytics:

- By viewing the public source code, or HTTP headers, of web pages. This can be automated and therefore is highly scalable. However, there is no way to determine if the tool used is the free or paid version.
- By surveying participants. This is limited by survey size and cannot be automated. Paid versus free usage can be determined.

89% of surveyed businesses use Google Analytics

A 2013 survey of 896 businesses (two-thirds of them from the UK) revealed that 89% are using Google for analytics (94% of agency clients). Of those using Google Analytics, 11% use the paid version, with 26% considering it.[3]

63% of Fortune 500 companies use Google Analytics

A 2013 study by Enor Inc. revealed that 63% of Fortune 500 companies use Google Analytics.[4]

Half of the web uses Google Analytics

A 2010 study by Metric Mail, who analyzed the pages of Alexa's top one million domains by searching for the Google Analytics tracking code text within pages, found Google Analytics was on 50% of the domains.[5]

The Paid Version

In September 2011, Google expanded its analytics offering with the launch of Google Analytics Premium—a paid version with a fixed fee of $150,000 per year. The Premium product is specifically targeted at the enterprise market. The key differences of Premium over its free sibling are

1 More horsepower for large data volumes

2 A service level agreement (SLA)

3 Installation, training, and support included with the fee

4 Additional features suitable for the enterprise environment

The Premium product is not a different kind of Google Analytics or a replacement for the free version. Rather, it is an extension of the free product to cater to the needs of enterprise clients. If you were to log in to Premium, you would be hard pressed to notice the difference from the free version. The user interface is virtually identical.

Why the Need for a Paid Version?

If you are an existing Google Analytics user, the main reason you would consider paying for Premium is data volume. If you regularly exceed 10 million data hits per month (roughly one million visits per month), then you need the extra horsepower of Google Analytics Premium. The free version is limited to 10 million data hits per month. This is written in the terms of service but is not restricted in the tool itself. That is, you can go over the limit and not lose data. But if you regularly exceed the limit, you can expect a call from Google politely asking you either to sample your data collection or upgrade.

If you are new to Google Analytics, item 2 may also be important to you. Large enterprises can have difficulty procuring a free product and would rather see a binding legal contract with commitments from Google with respect to data ownership and processing.

For a more detailed discussion of differences between the two products, see Chapter 2.

The Delicate Questions Answered

While I was working for Google, I spent a considerable amount of my time addressing the concerns I alluded to earlier in the chapter. The questions are still relevant today. Here is how I explained them.

Why Is Google Analytics Free?

The justification for a free model is this: Google is a media company; 91% of its revenue came from advertisers in 2013.[6] That is, advertisers pay to have their ads displayed next to Google search results. By the end of 2013, this was worth $50 billion per year.

So by removing any cost barrier for adopting a web measurement tool, Google was able to empower its customers with information about their advertising expenditures. The theory is that an informed advertiser is a happy advertiser. That's good for customer retention and for growing an advertiser's spend over time.

The justification for a paid model is simpler. Enterprises want face-to-face contact with their suppliers (account management), a contract to refer to if something should go wrong (SLA), more horsepower for large data sets (more than 10 million data hits per month), and features specifically relevant to them (such as BigQuery for database-type queries). See Chapter 2 for a more detailed discussion.

So there is no catch—it's a *freemium* model.

Can Google Be Trusted?

In other words, what will Google do with our data? That's a valid question to ask of any data behemoth. In my view, this boils down to two questions:

- Who—that is, what human beings—will see my data at Google?
- What non-human processing of my data takes place for Google's benefit?

The answer to the first question—who sees the data—is a small number of support staff and maintenance engineers who need to ensure the service is running day to day. Viewing client data is managed on a strict need-to-know basis. If you require someone to view your reports, then it is you (the account administrator) that initiates this with Google. In addition, you will need to explicitly grant report access to Google within your Google Analytics account settings.

In terms of non-human processing, clearly Google wants your data. The company makes a lot of money by selling highly targeted advertising and it is data that powers the targeting. Any data set that helps it do that is valuable to Google. Such data helps it improve its products—be it improving its advertising platforms (AdWords, AdSense) or improving tools such as Google Analytics itself. Like any data-savvy organization, Google values data.

Non-human processing concerns the analysis of vast quantities of data to understand trends. When you need to make business decisions based on billions of data points collected each day, looking at individual accounts would be both time-consuming and ineffective.

My view is that you should consider Google's use of your data in the same light that your website visitors (and customers in general) consider your use of their data. Trust is a key part of the relationship. Lose it, and not only do you lose the customer but that bad experience can also become viral on the web—and a great deal more expensive for you.[7]

Who Wouldn't Use Google Analytics?

Of course, working with Google may not suit everyone. With Google operating in so many areas now, it has quite a few competitors. If your business is in some way competing with Google, you may not wish to have your data collected and reported on by them.

Google does not build highly customized products. The secret to Google's success for all its products has been its scalability—growing to millions of accounts. If you require a highly customized data collecting and reporting platform, for example data retention beyond the 25 month limit (Google Analytics Premium has 36 months of data retention), Google Analytics is probably not right for you. That said, ensure you understand the capabilities of Google Analytics (chapters 2, 5, and 6).

Other than these two areas, I am struggling to find a reason why not to use Google Analytics. Having worked with Google products pretty much all of my professional life, and worked within the company for three years, I have seen both sides of the fence and consider myself independent (to the point of criticizing Google if I feel it justified). Recommending Google Analytics does not generate income for me. Working with great products does. I would not recommend them if I did not feel they met my high standards.

WHAT YOU CAN ACHIEVE

As I state in the introduction, my goal with this book is to get you to think in terms of *insights*—not data. An insight is knowledge you can relate to. It's a story or, more accurately, a coherent hypothesis (there are very few certainties when it comes to understanding anonymous data). The point of insights is to put you in the shoes of your visitors, so that you can understand their needs when they come to your website and view its content.

A visitor's needs depend on two key factors:

- Their expectations, discerned from how they got to the page—what search engine, campaign ads, or social conversation drove their decision to seek you out
- Their user experience—how easy it was to use your website, to navigate around, find information, engage with you (contact you, purchase, subscribe, give feedback)

It is your organization's ability to manage, analyze, and improve these two factors that determines your website's success (or not).

Clearly a set of data points, such as how many visitors come from Google, or what percentage went on to contact you/purchase/subscribe, sheds little light on those factors. On the other hand, the purpose of insights is to pull all of the relevant data points together to build a story of your visitors' journey and their experiences.

Insights come in two forms:

- Informative insights—those that help you understand your visitors and customers better, so that you can make informed decisions
- Action insights—those that require you to take action: reward a member of your team for a job well done; move your marketing budget to a better performing channel; remove friction from your website

Google Analytics doesn't provide insights by itself—no tool can. Producing insights requires an understanding of your business and its products; your value proposition; your website's content, engagement points, and processes; and of course your marketing plan. People, not machines, build insights. That said, people are inefficient without machines (tools like Google Analytics).

Google Analytics is a great tool for collecting and processing your data—that is, building reports. And it has over 100 default reports at hand. You can multiply that by a factor of 10 if you take into account all the customization and segmentation options. With ever-decreasing costs for processing, storage, and bandwidth, reports are cheap to produce.

Google Analytics has lots of reports. It has class-leading data visualization

features. You can import data, export data, query your data programmatically via its application programming interface (API), query your data with SQL (via Google's BigQuery app for very large data sets), and view demographic information about your visitors. You can animate, segment, customize, filter, and even watch your traffic performance in real time—a fascinating, almost musical experience as you watch the charts rise and fall and patterns come and go. There is more on features in Chapter 5.

The web metrics industry has for many years suffered from features wars—overcomplicating the product with tick boxes of features. That sort of worked when companies sent out RFPs to potential vendors. The RFP approach doesn't work anymore; that is the *big* change Google brought to this industry—a different approach centered around the value of data to an organization. Making Google Analytics free proved the point. And that is the point I am trying to make: improving your website comes down to understanding the two key factors I define that determine success— visitor expectations and the user experience. Google Analytics provides the data (and lots of it) that enables you to assess these.

By extracting data about visitors' expectations and their user experience, you can improve their experience. And that's what you can achieve with Google Analytics.

SPEAKING OF MEASURING SUCCESS …

When I hear this …	I reply with …
We need to see where our traffic is coming from. Tell us which channels are sending us visitors?	This is usually one of the first requests I receive when a client has a new Google Analytics setup. It's straightforward to answer, but a better request is "Show me which channels are producing my most valuable visitors."
Content is king on the web; search engines love it. Show me the most popular pages viewed.	A better request is "Show me the pages that our most valuable visitors viewed."
If we are smart about how we set up our offline marketing campaigns, can we see if these visitors are also influenced by our online marketing efforts?	Yes, and it's not rocket science. By that I mean that a good digital marketer can own this process. It is not an IT project.
Can we track the following: a visitor views an online ad on their laptop, views our website, visits our store, and checks our website again via their phone. Then they go home and purchases from an affiliate.	Yes. And if the visitor made the purchase within your store, you can also track that offline conversion in Google Analytics.
We have other tools in our organi zation that overlap with Google Analytics. They track some of the same things. However, when we compare these, the numbers are different. Which one is correct?	Avoid comparisons at all costs! They are fruitless and a waste of your time. When it comes to counting, there are differing techniques, methodologies, and definitions for the same thing. Therefore, use a centralized approach (one tool), when you need to compare metrics across the board. Google Analytics is intended for that purpose.

CHAPTER 1 REFERENCES

1 According to the e-tailing group's 12th Annual Merchant Survey of 2013, 46% of US merchants report a purchase conversion rate between 1.0% and 2.9%. www.e-tailing.com/content/wp-content/uploads/2013/04/pressrelease_merchantsurvey2013.pdf

2 Builtwith.com lists the number of websites running Google Analytics tracking as 25,218,196 as of the week beginning 21 July 2014: http://trends.builtwith.com/analytics

3 Econsultancy. "Online Measurement and Strategy Report 2013": http://econsultancy.com/reports/online-measurement-and-strategy-report

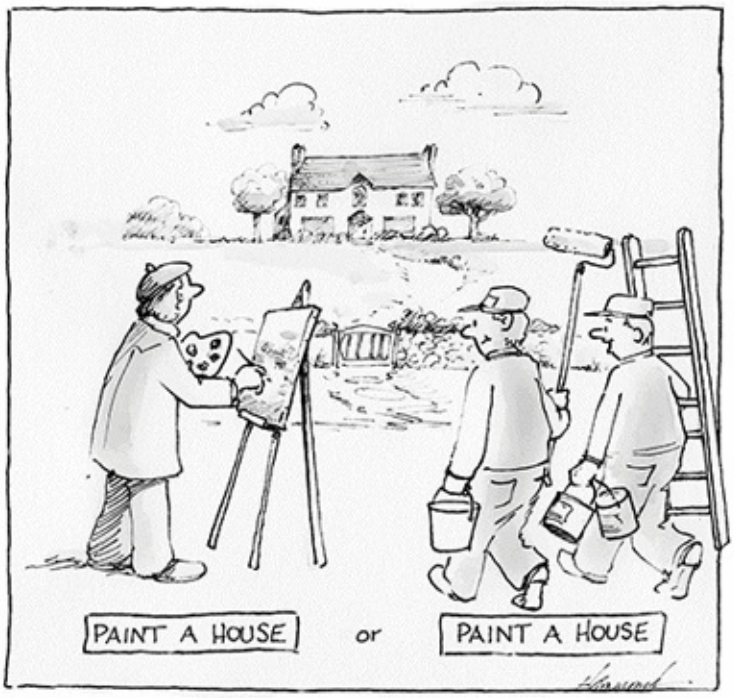
4 Enor Inc. study of Fortune 500 adoption of Google Analytics:www.e-nor.com/blog/google-analytics/google-analytics-solidifies-lead-in-fortune-500-adoption-in-2013

5 Metric Mail study of Alexa's Top 1 million domains: http://metricmail.tumblr.com/post/904126172/google-analytics-market-share

6 Google Q4 earning results: http://investor.google.com/earnings/2013/Q4_google_earnings.html

7 A great example of a bad experience going viral is Dave Carroll's "tribute" to United Airlines, who broke his Taylor guitar. With over 13 million YouTube views, Dave authored a book about his experience of corporate indifference that changed the perception of social media. www.davecarrollmusic.com/book

2 Choosing the Right Tool



Two versions of Google Analytics are available—the free version and Google Analytics Premium (the paid-for version). Which one is right for you? First off, let's get price out of the way. The standard free product is of course free to use, though you need to allow a budget for building a data strategy, implementation, and performing insights analysis. On the other hand, Google Analytics Premium is an annual contract, billable monthly, at a fixed cost of $150,000 per year.

Is that good value for money? For those websites with very high traffic volumes, Premium can represent great value for money when competitive tools bill by volume.With some other tools, the more data you have, the more you pay. Premium's price is fixed up to one *billion* data hits per month (approximately 100 million visitors per month). In addition, for budget planners a fixed-fee pricing model is advantageous for when your traffic can vary significantly, for example, by

season: you know exactly what your analytics bill for the coming year will be.

For websites with modest volume, spending $150,000 a year on data collection and reporting can look like a big expense. Therefore, the tool needs to be clearly justified within the organization.

In the middle are websites that have the necessary budget but need guidance on selecting the right tool for the job.

All users want to make an informed decision as to which is the right tool for their organization. No one wants to spend $150,000 a year on a product they could have gotten for free. Likewise, no one wishes to advocate a free product and risk not being taken seriously by colleagues due to underestimating the significance of the task it is required for.

THE VALUE PROPOSITION OF PREMIUM

From an analyst's point of view, the free and Premium products appear almost identical. In fact, if you were to log in to the user interface of each product, you would be hard pressed to tell the difference—either visually or by features. So why is there such a gulf in price? To answer that, consider the four cornerstones of the Premium value proposition, as shown in Figure 2.1: the contract, horsepower, features, and service.

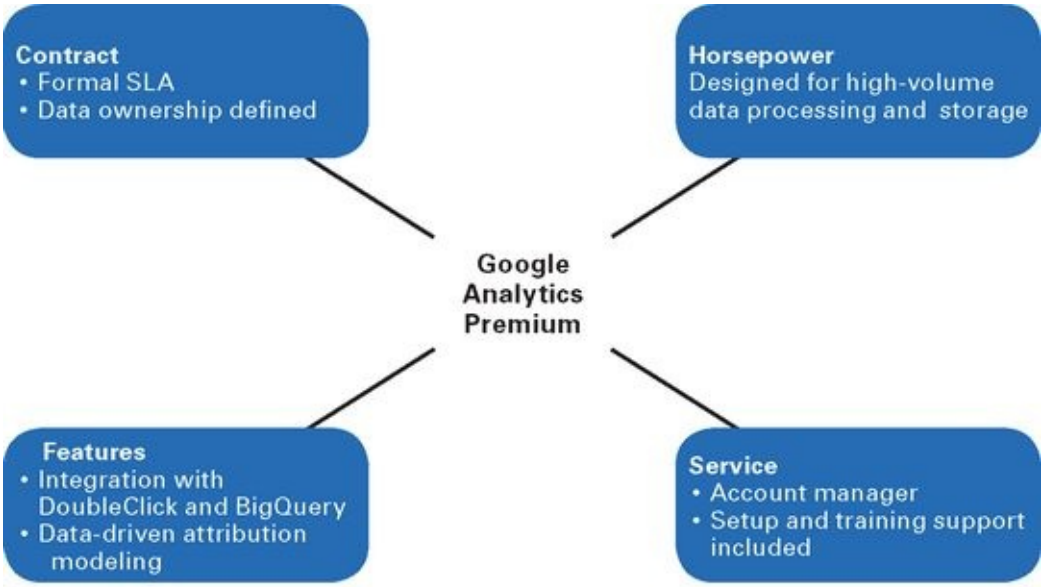


Figure 2.1 *The Google Analytics Premium value proposition*

The value proposition of Premium is not just about gaining access to extra features. Rather, Premium is tailored to the needs of the enterprise as a whole. For example, all users love the idea of free-to-use products. However, at the enterprise level, procurement departments have a hard time dealing with free. That's because many enterprises require clear answers to questions such as Who is responsible for this service? Who has access to it? What if it stops working? Who owns the data? Who is our account representative? These answers need to be formalized into a legally binding document before the product can be adopted as an official business tool within the organization. That formalization process isn't available with a free product. If this describes your organization, then Google Analytics Premium is the right product for you.

The Contract

Having a formal contract in place solidifies responsibilities and provides a direct point of contact for ensuring the service is running. And that means having an SLA in place. For Premium users, Google's standard SLA provides the following guarantees. These percentages are based on a calendar month:

- 99.9% guaranteed uptime for data collection
- 99.0% guaranteed availability of the user interface and reporting
- Data freshness guaranteed to be within 4 hours

(*Data freshness* means that the data in your reports is never more than 4 hours old.) If any of these levels is not maintained, you receive a credit against your next invoice.

As you would expect, Google provides no service-level guarantees for the free product. As any Google user knows, the reliability of Google products is extremely high and rarely an issue. In fact, I am only aware of two periods of unexpected Google Analytics outages that lasted more than a few hours since 2005. (Google is pretty good at running its networks!) That said, data freshness for the free product is delayed by the volume of data you have. Often freshness is within 4 hours, though it may be delayed by as much as 48 hours.

Horsepower

By *horsepower*, I am referring to the software capabilities with respect to its data processing and storage. Premium has more of this available. Having more horsepower means Premium can

- Process larger amounts of data and store it for longer than the free service.
- Maintain report freshness regardless of volume.
- Avoid report sampling.

Data Volume

Data volume only becomes an issue you need to consider if you have more than 10 million data hits per month. This is the limit set in the terms of service of the free product. That's quite a lot of data—approximately equivalent to 1 million visits to your website per month. However, popular B2C sites can have a great deal more than this. If that describes you, then you need the Premium product—or you will need to restrict the data hits you collect (see the sidebar "What Is a Data Hit?").

For Premium users, 1 billion data hits per month are included with your fixed fee. You can also purchase more—up to 20 billion data hits per month (data volumes can get very large for popular high-traffic websites). Note that the Premium limit is applied on the business entity—the Premium customer—not for each of your websites. If you have, for example, three web properties each receiving 100 million visits per month, one Premium account will cover your needs. That is, you do not require three accounts.

What Is a Data Hit?

From Google's perspective, a data hit is the data sent by each user interaction on your website. The only user interaction Google Analytics collects by default is the visitor's pageview. If you receive 1 million visits to your website each month and on average each visitor views 4 pages, Google Analytics will collect 4 million data hits.

Other common data hit types include events, transactions, and social interactions. If a visitor viewed 8 pages on your site (8 pageviews), downloaded a PDF brochure (1 event), made a purchase (1 transaction of 2 items), and then clicked on your Facebook "Like" button, that would be a total of 13 data hits.

When estimating data requirements, I allow for 10 data hits per visitor. In other words, 1 million visitors per month will generate approximately 10 million data hits.

Restricting Your Data Hits

You are required to restrict your data hits if you are regularly exceeding the 10 million data hit limit of the free Google Analytics product. If, for example, your website receives 50 million data hits per month, you need to restrict your data collection at 20% in order to remain within the terms of service for the free product.

Restricting your data collection can be a strategic decision (you simply do not collect parts of the visitor

journey in the first place, such as not tracking visits to your blog area), or a setting you make within your tracking code on your pages to randomly exclude visitor sessions in your tracking. The latter is preferable as it is statistically safer.

Report Freshness

Report freshness becomes slower the larger the volume of data Google Analytics has to process. However, Premium guarantees this will never go above 4 hours. If you are using Google Analytics for free and have small to medium website traffic, you will see a similar time for freshness. However, the delay will increase with data volume. If your website receives over 1 million data hits in a single day, your reports will only be refreshed once for that day in the free version. That is, you will have no intraday processing.

Report Sampling

Most Google Analytics reports are not sampled (free or Premium version). However, as you drill down into your data, the data associations become more complex. As with all software products, as complexity grows the performance deteriorates. To avoid this impact on the service, Google Analytics introduces sampling limits.

For the free version, sampling kicks in when the number of visitor sessions in the *analysis pot* exceeds 500,000. Typically you will see this if you have a large volume of traffic or you are viewing data over a long date range, such as year-on-year comparisons.

For Premium, the sampling limit is 20 times higher, at 10 million sessions, though this value is continually being revised upward. Reports can also be downloaded unsampled up to a limit of 100 million sessions.

Sampling Explained

By sampling visitor sessions at random, Google Analytics reduces its processing load. For example, if the analysis pot you are investigating includes 2 million sessions, the free Google Analytics will sample 1 in 4 of those sessions at random to bring the pot size down within its limit. After processing the sample, Google Analytics scales the results back up by multiplying by 4. This is a standard statistical method when dealing with large volumes of data. A smaller representative subset of data is used to estimate the total values.

Although the scaled numbers are statistically accurate, problems occur if you further analyze the sampled report, such as when you drill down within your sampled data to examine conversion rates or revenue numbers. As these are an even smaller subset of the sampled data, the error bars of the scaling become significant. Sampling is discussed in further detail in Chapter 5.

Features

There are not a great deal of feature differences between free Google Analytics and the Premium version. This is deliberate. After all, powerful features are what *all* analysts require—the direction of Premium is not about pro versus amateur features! However, a small number of enterprise-class features are tailored for the Premium product:

- Integration with DoubleClick
- Google Cloud storage and BigQuery
- 200 custom metrics and 200 custom dimensions
- Data-driven attribution modeling function

Integration with DoubleClick

DoubleClick is Google's ad management software that manages display advertising—essentially banner ads. Content-rich websites that sell ad space on their pages can use DoubleClick to power the display of ads. The same product also allows advertisers and agencies to manage their campaigns. Well-known brands—enterprises—are the typical users of online banner advertising, and DoubleClick had 18% of the US display ad market in 2013.[1]

Premium integrates with DoubleClick so that advertisers can view ad costs, impressions, and click-throughs and view-through traffic (those users that viewed your banner ad but did not click through). Because DoubleClick uses a third-party cookie to track ads displayed across the web, it is also able to gather demographic information of website visitors that see those ads. This non–personally identifiable information includes gender, age range, and interest category data—sports, tech, food and drink, travel, and so forth—and their subcategories.

The free version of Google Analytics does not integrate with DoubleClick.

Google AdWords Integration

Contextual text ads that appear next to search results on Google are powered by Google AdWords. Both Google Analytics free and Premium versions integrate with AdWords, and the relevant reports are identical for both products.

Google Cloud Storage and BigQuery

Google Cloud storage gives you the option to export your data and query it using a variant of Structured Query Language (SQL)—a standardized way to query data that is familiar to all analysts. Google's variant is called BigQuery. It's a superfast query engine for massive data sets—data sets too large for standard relational databases.[2]

The result is that BigQuery allows you to build complex, specific queries that go

way beyond the capabilities of what can be achieved within a graphical user interface. The free version of Google Analytics does not integrate with Google Cloud or BigQuery.

In addition, Premium guarantees to store your data for a period of 36 months; the free version is limited to 25 months. In both cases Google has never actually deleted any data (I have reports going back to 2005). However, it reserves the right to do so if those time limits are reached.

More Custom Metrics and Dimensions

The free version of Google Analytics has the ability to set 20 custom metrics and 20 custom dimensions. Think of these as labels assigned to your visitors during their sessions. A typical custom dimension I often use is *customer* versus *noncustomer*. The visitor is labeled as a customer if they have purchased from you. By using a cookie, the label can be set to stay with the visitor on subsequent visits. Knowing this visitor label lets you separately assess the visit behavior of your customers versus noncustomers. The behavior of these two visitor types (segments) on your site can be quite different and informative. Other potential custom dimensions include *subscriber*, *logged-in*, *commenter*, *engager*, and *social sharer*.

Custom metrics and dimensions also provide flexibility—for example, to integrate with your CRM system or your marketing automation system. If these tools rate or score your visitor engagements, you can pass this information back into Google Analytics as custom metric or dimension, such as *top-tier customer* (dimension), *propensity to become a customer = 58%* (metric), *noncustomer but spoke with sales rep (dimension)*, and so forth.

Premium works in exactly the same way; the difference is you get ten times as many to experiment with.

Metrics versus Dimensions

Google Analytics reports consist of two different types of information—metrics and dimensions.

A metric is a number, for example, the number of visitors to your website, the number of conversions from a campaign, the amount of revenue gained, or new leads generated.

A dimension is textual information, for example, the list of your top-performing pages, the most effective campaign names, or your best-selling products.

Data-Driven Attribution Modeling

Standard attribution modeling techniques allow you to attribute the revenue you make (or value from lead generation) back to the referring websites who sent you your visitors. This is a common technique if affiliates are an important part of your marketing efforts. However, it can be applied to all your website referrers. Consider Figure 2.2.

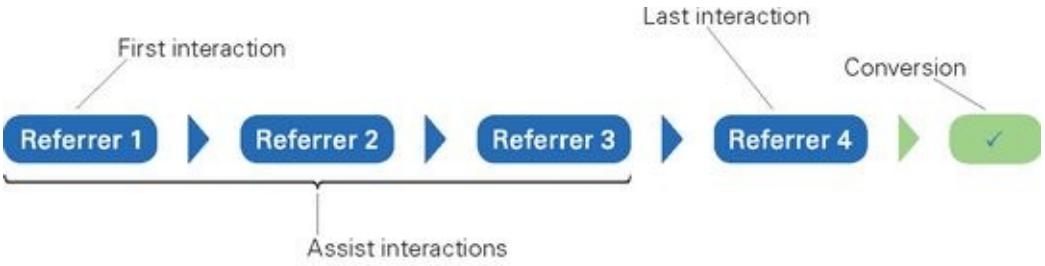


Figure 2.2 *Anatomy of a conversion path. In this example,* path length *(number of interactions to conversion) = 4.*

Figure 2.2 schematically shows a visitor to your site who made 4 visits before finally converting (becoming a customer, subscriber, or lead). Without attribution modeling, all referrers are considered equal. If the referrers for this path were a Google search, followed by a social media conversation, followed by a banner ad click-through, followed by a click from your affiliate, each of those referring sites would be considered to have an equal importance. Attribution modeling allows you to give a different weighting (a different proportion of the resulting revenue) to your referrals.

Both free and Premium have the same attribution modeling techniques to choose from—you can even customize your own. However, Premium has one extra—data-driven. With this feature, rather than your having to manually compare and choose from different models (a complex process fraught with caveats that frustrates even experienced analysts), data-driven attribution modeling does the work for you. The model *automatically* distributes credit across marketing channels. Google achieves this by using sophisticated statistical techniques to calculate what the weighting of each referral should be—based on analysis of paths that convert as well as those that do not. Attribution modeling is discussed in more detail in Chapter 6 (in ebook 2).

Service

Historically, enterprises have struggled with web analytics. Typically they paid out a great deal of money to get the latest tool with the 10,000 or so features they were told they needed (a slight exaggeration, perhaps). Once the tool was in place, it seems no one informed them they had to regularly invest further resources to actually build a data strategy, configure the tool accordingly, investigate the data, understand it, go back and adjust the setup, further investigate, communicate to the business, and so on.

My friend and former work colleague at Google, Avinash Kaushik,[3] came to prominence around 2004 by describing this with his 10/90 rule—meaning organizations were investing 90% of their money in having the tool and only 10% in using it to make smart decisions with it. Lambasting this shortsighted approach, he evangelized passionately to reverse that ratio. (With Google launching their free web analytics tool in 2005, hiring such a smart and vocal advocate was a no-brainer.) Avinash's point is that it's what you do with your analytics tools that matters, not the tool itself.

When Premium launched in 2011 with its fixed-fee pricing model, a key element was service, including help with building a data strategy, getting the tool set up to reflect this, and training your team to understand how to use the product. It is not an unlimited supply of resource, but it will ensure you have a best-practice setup that meets the expectations of your business and that your organization gains the internal knowledge to use it. Service can be provided by Google Analytics Premium authorized resellers (www.google.com/analytics/premium/partners.html), or direct from Google itself.

As you will have guessed, no service is included with the free Google Analytics product. However, a specialist partner network exists of Google Analytics Certified Partners that can tailor a service offering around the free product (www.google.com/analytics/partners).

Google Analytics Certified Partners (GACPs)

Google has an official global network of over 200 certified Google Analytics partners, known as the GACP network (www.google.com/analytics/partners). These companies help you with service requirements such as data strategy, implementation, training, and analysis. Regardless of which product you use, you should at least seek out a GACP for advice.

Google Analytics Premium authorized resellers are a subset of the GACP network.

DETERMINING THE SELECTION CRITERIA

The Econsultancy–Lynchpin "Online Measurement and Strategy Report 2013"[4] posed two important questions to its surveyed audience:

- For Premium users: What are the key reasons for using Google Analytics Premium?
- For non-Premium users: What is the main reason your organization doesn't use Google Analytics Premium?

The responses to these two survey questions are shown in Figures 2.3 and 2.4. The lack of importance given to the sampling of data is surprising to me—see Figure 2.3. It indicates that good governance (data ownership, an SLA, and included support) is considered more important at the enterprise level than having more processing horsepower—although maybe these Premium users do not have high traffic volumes and therefore sampling is not an issue. Whatever the situation, I predict that processing power will become a more important part of the Premium offering as data volumes grow beyond websites to mobile app usage and other Internet-connected devices, such as store checkout tills.

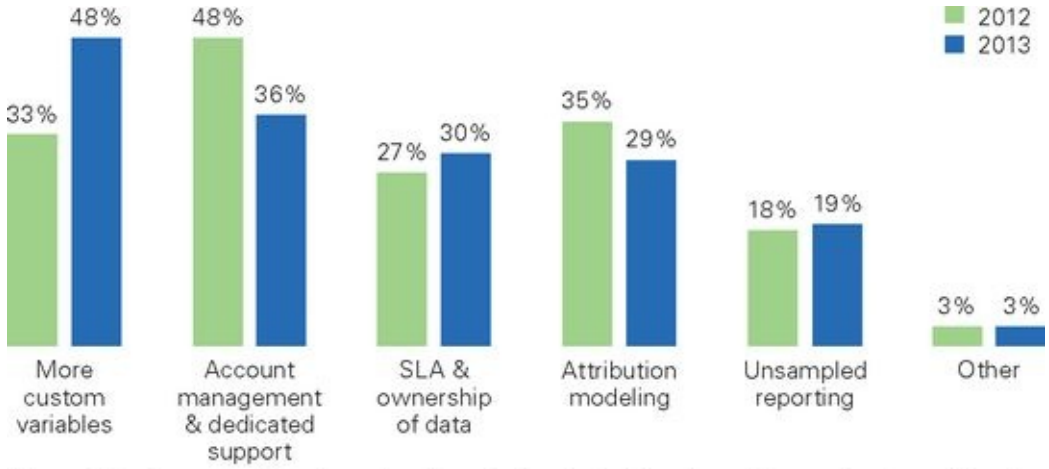


Figure 2.3 *Reasons given for using Google Analytics Premium. Respondents could select up to two options.*

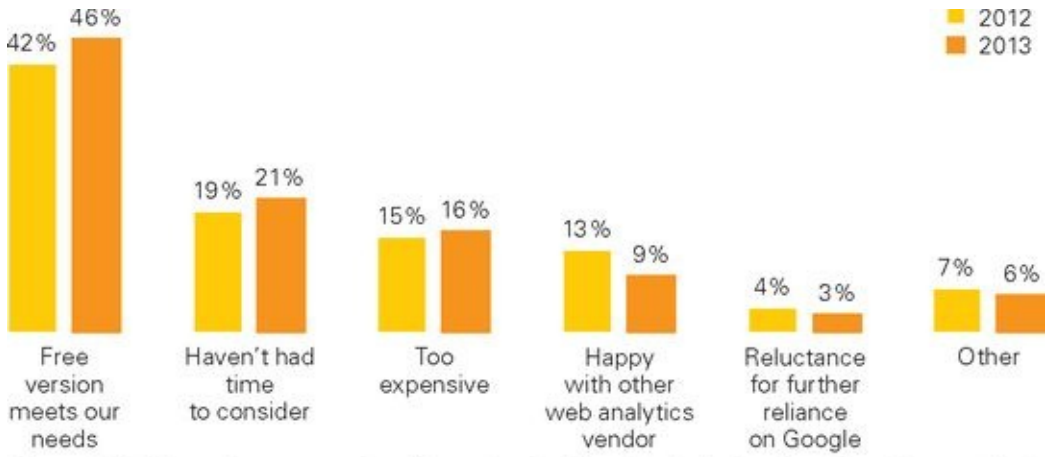


Figure 2.4 *The main reason given for not using Google Analytics Premium. Respondents could select up to two options.*

Figure 2.4 shows how good the free product is—nearly half of all respondents said it met their analytics needs.

For analytics needs, there are four types of potential Google Analytics users:

- Existing Google Analytics users currently using the free product considering the upgrade to Premium
- Existing Google Analytics users currently using the free product and not considering the upgrade to Premium
- Existing analytics users—users of another analytics vendor (or internal proprietary tool), looking to switch to using Google Analytics
- New analytics users—not using an existing analytics tool

Regardless of which type of potential user you are, the flow diagram of Figure 2.5 guides you through the key decisions you need to make. The following sections consider each of the potential user types in terms of defining the criteria to determine the right tool for the job.

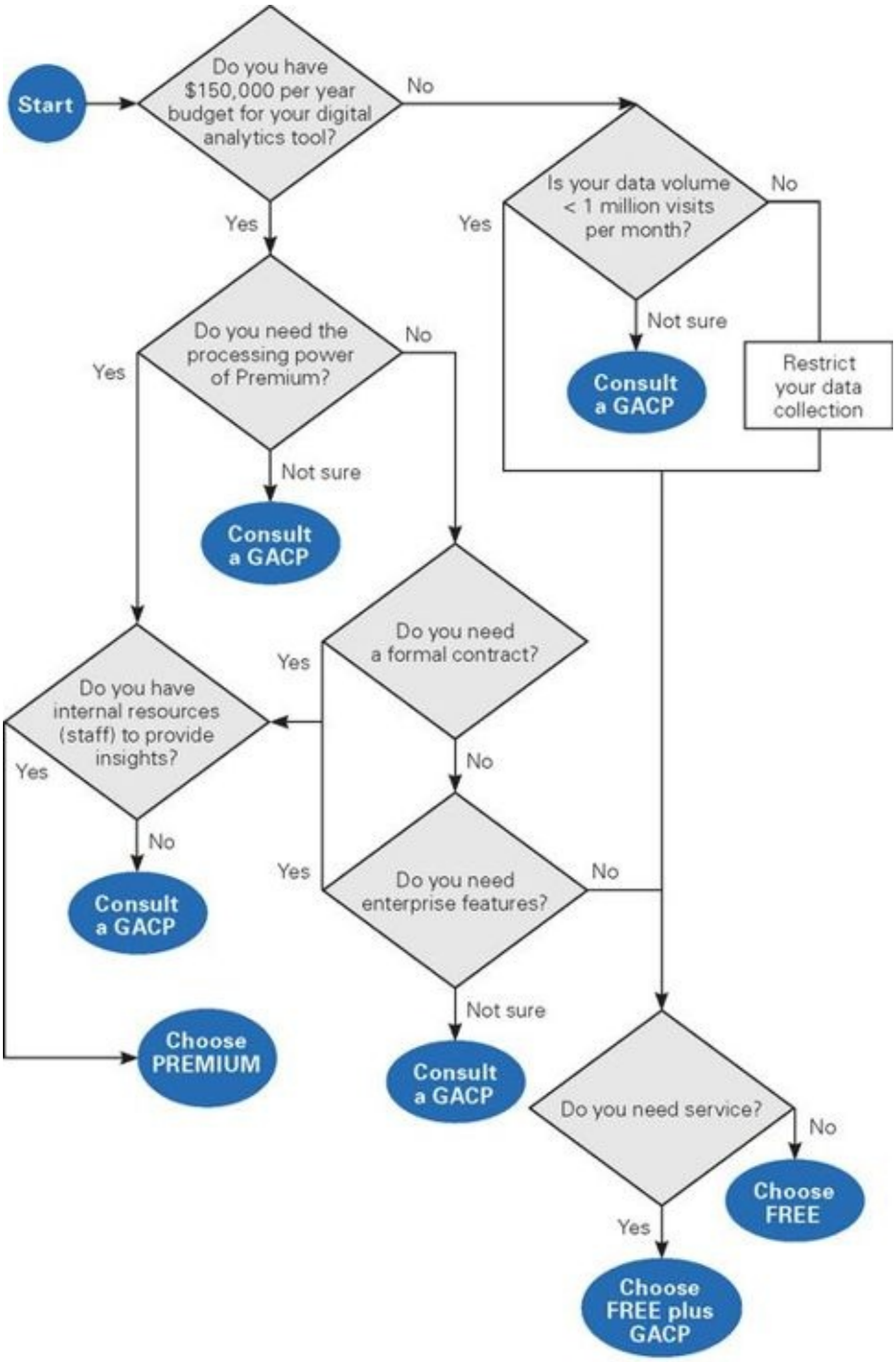


Figure 2.5 *Choosing the right tool*

What If My Budget Is Limited?

If $150,000 per year is not available for data insights, your options are restricted to using the free version. The free version is a great product in its own right—I have used it with numerous global organizations and helped them achieve real and significant gains. Using the free tool is about managing expectations—an important one being data volume. You will need to restrict this if it exceeds 10 million hits per month.

But bear in mind that a key necessity for using any analytics tool is ensuring you have a budget for maintenance (data quality) and building insights for your organization (not reports). Therefore, if your budget is limited, use the free tool and invest in these. Without good quality insights, your data project really is dead in the water. Consult the flow diagram in Figure 2.5 for guidance.

Existing Google Analytics users considering the upgrade to Premium If you

are currently using the free version of Google Analytics, you are familiar with its value proposition—it has class-leading features (see Chapter 5 for examples), it is intuitive to use and comparatively straightforward to implement, it has a large user base spanning the entire spectrum of organizational sizes, it is used globally by more than 25 million websites, it has an active community, and of course it's free. Google Analytics may not yet be an integral part of your business strategy, but it can be given the necessary investment. That investment means people. It includes training, education, hiring, even external help—and a great deal of time and energy to pull all of that together to drive insights. Even then, your insights need to be trusted and therefore adopted by your organization so that you can move forward. None of these depend on what tool you are using.

If this describes your situation, my advice for you is to focus on getting the most out of your existing free Google Analytics setup. For you, the only Premium criterion you should keep an eye on is your data volume. As long as you do not exceed 10 million data hits per month, you are unlikely to need Premium. Instead, invest your budget in improving your setup (data quality) and using it (gaining insights). Once you have maxed out on insights, you can circle back and consider what additional capabilities Premium can offer.

Existing Google Analytics users not considering the upgrade to Premium The vast majority of existing Google Analytics users will not be suitable for Premium. As shown in Figure 2.4, nearly half of all potential Premium users say that the free tool meets their needs. Usually this is because data volume is not a problem. That is, you receive fewer than 10 million data hits per month. However, at the enterprise level it is worth considering the full value proposition of Premium, as shown in Figure 2.1. If none of those are required for your organization, then great, you have just made an informed decision and saved your department $150,000 per year. However, invest that saving in building your processes—that is, ensuring the data stream is of continuous good quality—and then use it to drive insights and improvement (as described in Chapter 3).

Existing analytics users Many enterprises already spend $150,000, and often more, on analytics tools and processes. For these organizations, the Google Analytics value proposition is the same as for the free product—ease of setup, ease of use, class-leading data visualization, and a large user community sharing tips, tricks, and best practices. That is, you are considering switching to Premium because it is recognized as a good product (a great product in my view, packed full of innovation). If you can save budget by switching, that's a bonus. However, price is not your primary motivator. You want a tool whose vision and approach are helping you gain insights from your data—not bombarding you with features, complexity, or extra billable days of consulting.

If you are already investing Premium-size budgets in your digital analytics tools but are looking at Google Analytics as an alternative, the decision should be straightforward. All four areas of the Premium value proposition will be important

to you (Figure 2.1). When comparing the price with your existing tool, make sure you factor in the service part that Premium includes. I have seen many white elephant analytics tools where the client's budget has been exhausted on license fees alone, without considering the other areas that need investment in order to make it all work. The Premium service avoids that problem.

New analytics users If you are considering Google Analytics for the first time and have no legacy system in place, then getting the price right for your project will be a key initial question. Let's face it. No one wants to spend $150,000 per year on a product they could have gotten for free! It's also not about organizational size. For example, there are numerous multinational companies using the free version of Google Analytics—not because they can't afford a paid tool but because it does the job they need done.

Having an idea of what data volumes your website will generate is your first question. If you expect more than one million visitors per month, then Premium needs to be your default consideration. If you do not know what traffic volume may be generated, use your digital marketing budget as a proxy. For example, if your digital marketing budget is above $1 million per month, you are likely to generate more than 1 million visits, you are likely to be using Google's display advertising network, and you will want to dissect the results of all your marketing efforts rigorously. For any of these you should explore Premium in detail.

If neither of these applies to you, then the next question to consider is whether you have a need for a formal contract and SLA. If you are comfortable without having these in place, then you can safely start off your analytics investment with the free product. As your knowledge grows (typically over the next year or so), you can consider the value of upgrading to Premium again.

A Painless Upgrade Process

If you are already a free Google Analytics user, then the upgrade to Premium consists of Google flipping a switch once your contract is in place. No changes to your site or its tracking code are required, and you can view all your historical data after the switch.

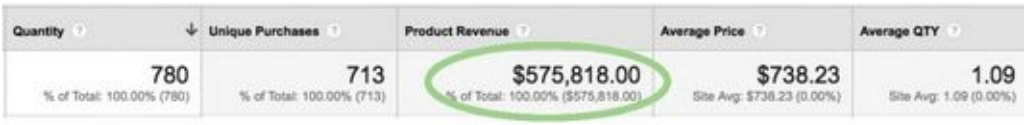
JUSTIFYING THE BUDGET

At the enterprise level, a budget of this size can always be made available. However, justifying the idea that budget should be spent on your team and its tools is what determines if you actually get it. So how can we justify $150,000 per year on a tool, and what else do you need to allow for to make it a success?

The answers to these questions are *value* and *resources*, respectively. That is, calculate what your website is worth to your organization (its dollar value), then invest in the resources to measure and improve it (its performance and value).

Website Value: What Is Your Site Worth?

Calculating the value of your website is straightforward when you are a transactional site. You look at the total amount of revenue collected in the relevant report. Figure 2.6 is an example taken directly from a Google Analytics report for a travel website. The value of the site is $576,000 for the month.



Quantity ↓	Unique Purchases	Product Revenue	Average Price	Average QTY
780 % of Total: 100.00% (780)	713 % of Total: 100.00% (713)	$575,818.00 % of Total: 100.00% ($575,818.00)	$738.23 Site Avg: $738.23 (0.00%)	1.09 Site Avg: 1.09 (0.00%)

Figure 2.6 *E-commerce performance data is readily available in Google Analytics.*

In addition, you have reports available where you can quickly ascertain your conversion rate, the value of each page on your website, and the value of each visitor—broken down by marketing channel and your specific campaigns. Both free and Premium versions automatically do this for you (it's a great feature of the product).

If you are a non-transactional website (the vast majority of enterprise websites are not e-commerce), then value is much harder to determine. In fact, I rarely see the calculation even attempted. However, without knowing the dollar value of your website, you are missing out on the big picture of what your website is capable of.

Monetizing your non-transactional site is critical to your success. Without it, there is a serious danger your website is just a pet project—nice to have but not treated as a serious part of the business. Hence $150K per year for a measurement tool is never going to see the light of day in your organization, and rightly so. As Jim Sterne, founder of the eMetrics Summit series[5] said, "You need to calculate the return on investment for measuring your return on investment."

Monetizing a non-commerce website is not rocket science but does require time and energy to logically think through your website's goals and assign values to them. For example, if your website generates sales inquiry leads and you know what percentage of them result in a sale, and you also know your average order value, then you can assign a value to your website leads.

Here's an example:

- Your website receives 250,000 visitors per month.
- Your lead generation form (new sales inquiry) has an industry average conversion rate of 3.0%. That is 7,500 sales leads per month.
- Your sales team is able to convert 1 in 10 leads to a customer (750).
- Your average order value, which happens away from your website, is $1,000.

Calculating the value backwards results in your lead generator form's being worth $100 per lead. That is a total of $750,000 per month (the value of your website), or $3 per visitor.

The calculations can get a great deal more involved when you have multiple

ways a visitor can become a lead—such as file downloads, newsletter subscriptions, email links, click-to-call buttons, and video demonstrations. But the calculations are no more complicated than this.

Taking this calculation a step further, not all website revenue is profit. Assuming you have a 50% profit margin, your website is generating $375,000 per month of gross profit (sales minus cost of sales).

Calculating Your Website Conversion Rate

Google Analytics calculates your website conversion rate for you automatically. It is defined as

$$\frac{\textit{number of visits that convert*}}{\textit{total number of visits}}$$

* Purchase or complete a lead generation form

If you have multiple conversion points (for example, a purchase and a lead generation form), you can select the individual conversion rates within your reports.

When Website Value Increases

Now let's suppose that with a good understanding of your digital traffic you can increase your lead generation (or e-commerce) conversion rate by a conservative 5%—that is, from 3.0% to 3.15%. Using $750,000 per month as our benchmark, the additional improvement in the conversion rate yields an extra $37,500 per month in revenue, or $18,750 per month in gross profit. Annualized, this will pay for your Premium license with $75,000 left over to invest in insights—that is, your staff!

I consider this the tipping point. If you value your website at approximately $9–$10 million per year ($750,000 × 12), you can justify a $150,000-per-year investment in an enterprise-class web analytic tool. That is based purely on being able to deliver a 5% improvement in sales or leads, or alternatively being 5% more efficient with your current digital marketing budget. Most probably it will be a combination of both.

My purpose with this section is to provide you with guidance on how to estimate the value of your website so you can make an informed decision on whether free or Premium is the right tool for you. Of course, the real value of your data is more involved than the calculation I have presented. An extreme example of data value is Twitter—valued at $18 billion in 2013, yet its losses for the third quarter were $64.6 million.[6] Twitter's value comes from the fact that it has data (tweet patterns, geographic and biographic data) on over 250 million active users. Clearly a different type of calculation is required when your data has intrinsic value.

Justification Rule #1—the Tipping Point

If your website is not contributing at least $10 million per year to your business, do not pay for Google Analytics Premium. Instead, invest in the resources required to get the most out of the free version.

Resources: Do You Have Them?

If analysis of your digital data is going to be someone's full-time job—that is, working on bringing insights to the business on a day-to-day basis—then consider Premium the right tool (even more so if there is a team of people involved). On the other hand, if the investment in analysis is only part-time, say only one day per week, I recommend the free product.

Why is staffing so important?

Spending only one to two days a week analyzing your website reports will not allow you to explore the capabilities of the Premium product in any depth. You would be constantly scratching the surface without the time to go deep. Enterprise-level analysis is a full-time job for at least one person. Incidentally, this is the same for the free product, but at least it has no cost.

Of course, you could (and should) hire a consulting expert to do the deep-dive stuff for you. However, that external person or organization still needs to be integrated within your business, or they become a data silo—capable of little or no impact—in which case you need the resources to manage and integrate them properly. Chapter 8 (in ebook 2) discusses resource and team building in more detail.

Justification Rule #2—People Are More Important than Tools

If you cannot commit one full-time person to work with your data insights (even if you hire a third-party expert to help), do not pay for Google Analytics Premium.

GAINING EXECUTIVE BUY-IN

Part of my goal with this book is to get you to move away from thinking about your digital performance in terms of data. That's a dichotomy for me—data of course are your building blocks and a great deal of this book (and my life!) concerns understanding its collection and quality. However, a data report is a snapshot—a status update, if you will—that allows you to see where your website is at that particular time. That is useful initially, but *insights* are where you need to get to rapidly if you want your data to help improve your business. Otherwise, your web analytics project just remains a cost to the business with no return on investment.

An insight is *knowledge*. It's not a data report. An insight tells a story of what is happening on your website and is therefore derived from your reports. It takes the form of a hypothesis—what you think is happening and why—built from cross-referencing your report data, looking for correlations, segmenting, visualizing, trending, and so on. This is followed by an action plan—what you are going to do about it. The process requires intelligent people with the propensity to change—that is, to not be afraid to take a calculated risk (based on good data) to change and experiment, with the objective of improving things and gaining knowledge. This is the role of your analyst team.

Analysts thrive on data; executives need insights.

Regardless of which product is right for your organization, you need to set aside a budget for insights—effectively a budget for building your team. Your team could be one or more people working as a bridge between marketing, web development, app development, and sales. Or, if your organization is still in the early stages of analytics maturity, your insights team could be a third-party GACP agency. Often it is a combination of both, where in-house immediacy and business knowledge work alongside a consultancy with a breadth of digital analytics experience (see Chapter 8, in ebook 2). That way, your data can mature into insights.

When you bring insights (not data) to the other senior decision makers of your organization, you will gain the buy-in to grow your team and invest in the tools you need. Every manager wants to base decisions on good data—I have never found resistance to the principle of measurement itself. However, providing insights, such as savings made or opportunities identified or grasped, will give you more control over the level of investment required to go further. What executive wouldn't want more of those?

SPEAKING OF CHOOSING THE RIGHT TOOL …

When I hear this …	I reply with …
We are a startup and estimate our website will be worth about $1 million a year to us. Should we be using Premium?	In terms of revenue numbers alone, my answer is no, use the free product, and invest in insights first. That is, hire someone or use a third party such as a GACP. You will see a much greater return on this than by spending your budget on a tool. Review the decision each year—upgrading is straightforward and painless.
We are a web portal and our free version of Google Analytics shows 90 million visits per month. Our margins are tight and we do not wish to invest in Premium. What are our options?	You are currently in breach of the Google Analytics terms of service. You need to restrict the data you send to Google by either not tracking certain interactions (for example, turn off event tracking) or sampling the data you collect (for example, only track 10% of sessions).
We track multiple websites around the globe, mobile apps, and point-of-sale transactions in stores. In short, this is more than 1 billion data hits per month. Can this all be tracked together so we can compare apples with apples?	Yes, any Internet-enabled device is capable of being tracked by Google Analytics (both free and Premium versions). For such high data volumes, you will need Premium. You can extend your fixed-fee contract with tiered pricing up to 20 billion data hits per month.
We have the budget to either hire a digital analyst and use the free product, or use Premium—but not both. What is the best option?	Always, always hire good people ahead of investing in a tool. This does not need to be a full-time position. For example, you can pay as you go with the services of a GACP. If or when you have exhausted all the possibilities of the free version (that can take some time!), look back and consider what Premium can offer you in addition. Review Chapter 8 (in ebook 2).
We have a large volume of data but no internal resources to help us understand it. In short, we do not even understand how to use the free version. Hence the CEO will not invest in Premium, and we are breaking the current Google Analytics terms of service. What shall we do?	If you are still trying to come to grips with your digital measurement approach, stick with the free product and restrict your data collection to below the 10 million hits per month threshold. Invest your budget in insights—either hiring someone or using a third-party GACP, or both. When you have grown your analytics maturity sufficiently (this typically takes a year or two), revisit your tool requirements.

CHAPTER 2 REFERENCES

1 www.bloomberg.com/news/2013-02-28/facebook-buys-microsoft-atlas-ad-placement-tool-in-google-combat.html

2 https://developers.google.com/bigquery

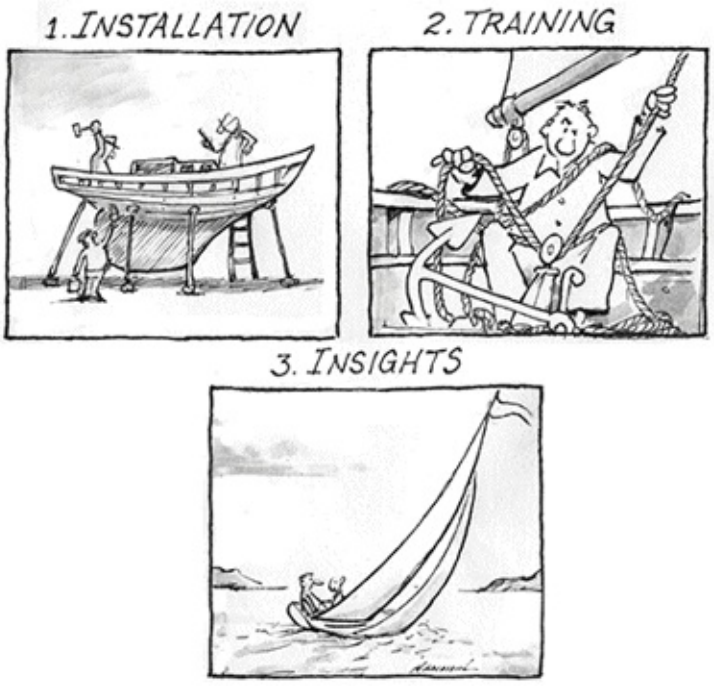
3 Avinash Kaushik is probably the most popular of all web metrics gurus. As *the* Digital Marketing Evangelist at Google, he has authored two best-selling books on web analytics and founded the popular blog Occam's Razor http://google.com/+avinash. Listening to Avinash speak at conferences is well worth the money.

4 Econsultancy and Lynchpin. "Online Measurement and Strategy Report 2013": http://econsultancy.com/reports/online-measurement-and-strategy-report

5 The eMetrics Summit series: www.emetrics.org

6 http://www.bbc.co.uk/news/business-24819238

3 Setting Expectations and Building the Process



Although expectations and processes are by nature specific to each organization, there is a great deal of commonality. Regardless of organization size, business sector, market (geo-location), or analytics tools used, the steps required to start off a digital measurement project—and what it takes to make it all work—are remarkably similar. It's how you achieve them that is specific to your organization.

Chapter 1 (Figure 1.3) outlines my view of the digital measurement world, using Google Analytics as your unified web measurement platform. That is a realistic ambition, though it does take time, energy, and money to achieve. Figure 3.1 shows the four required phases of implementing a digital data strategy within your organization.

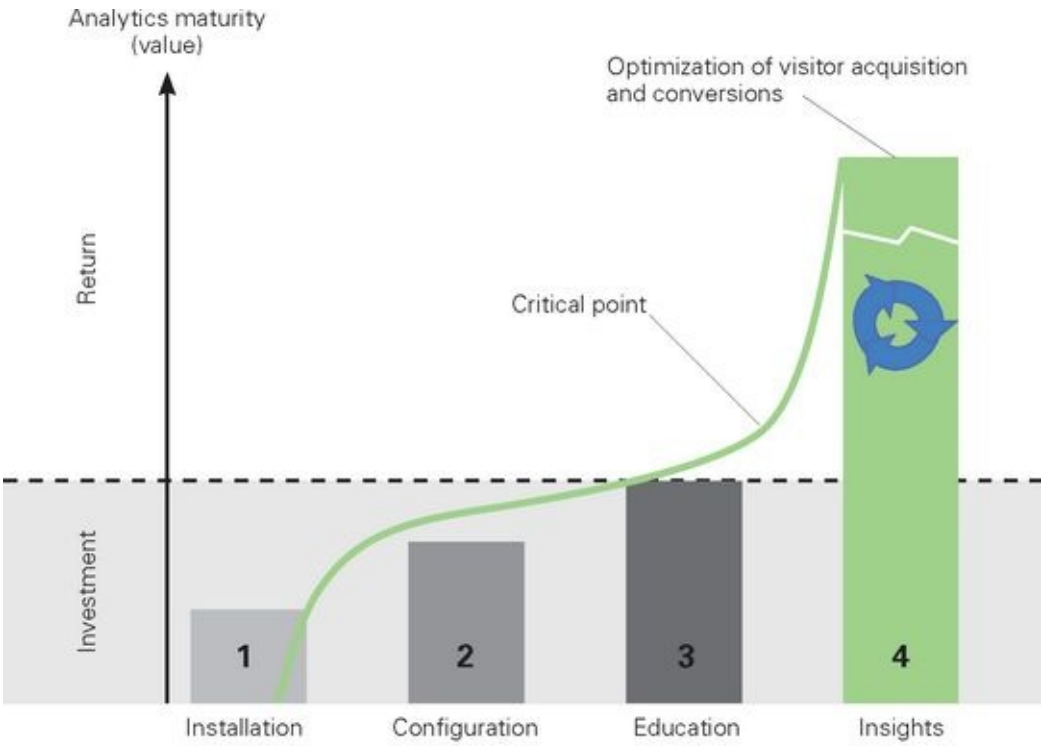


Figure 3.1 *The four-phase process*

Applied in order, these are

1 Installation—getting the data in by installing Google Analytics. Collecting data from your website, mobile apps, checkout tills, or any Internet-connected device that is relevant to your customers.

2 Configuration—structuring the data so it makes sense—for example, defining filters, segmentation, custom reports, or campaign tracking.

3 Education—ensuring your team understands how to use Google Analytics and interpret the reports.

4 Insights—studying the data to glean information and ultimately gain knowledge to drive informed change of your organization's digital strategy.

The vertical axis of Figure 3.1 is a representation of what your data gives back to your organization. It can be monetary value or a perceived value—that is, your organization's analytics maturity (its ability to use data effectively.[1] At the lower extreme of Figure 3.1, the data has very little value to you—it is not trusted to be used for making important decisions. Conversely, at the higher extreme the data is relied upon to drive the business forward.

The first three items (installation, configuration, education) are a cost to your organization. You need to invest in order to make them happen. Hence the value or maturity during these phases remains low but is growing rapidly. The phases must be applied in order—installation comes first—though the phases do overlap. For example, once data starts to come in, you can commence with configuring (structuring) the data—it is not necessary to complete one phase before commencing the next. However, do not change the order of the phases. For example, training people before you have structured your data is a sure way to

waste valuable training days—people cannot focus on, or maintain interest in, data that makes no sense to them.

The fourth phase, insights, is where your organization needs to be in order for you to see a return on your analytics investment. That is the phase where you gain back more than you put in. This can be because your data has helped you improve your website and therefore convert a greater percentage of visitors into customers. Or it can be because your data has enabled you to better focus your marketing budgets—that is, gain more customers for the same marketing spend. Both are good news.

The change in the value of your data when you reach phase 4 can be dramatic. For example, the graph shown in Figure 1.2 shows a return that is more than 100 times greater than the total analytics investment! This critical point is reached if and when you are willing to take on board change—such as changing your website architecture or its content, or changing your marketing approach. That sounds straightforward and quite obvious. However, I have found that an organization's ability to change can be a major hurdle to moving significantly beyond the dotted "break-even" line of Figure 3.1.

Repeating the Process

The last phase of Figure 3.1 shows a cyclical arrow. This is to indicate that all the phases (1–4) should be repeated periodically. For example, over time the data quality of your Google Analytics implementation will deteriorate. This is because of web content updates, mobile app updates, developers forgetting to add the tracking code, the tracking code no longer reflecting the actions your visitors take, technology changes, people using the web in new and different ways, and so forth. It is therefore important to continually assess all four phases of the process and adjust accordingly.

The optimal reassessment frequency depends on how often you change your digital content. I recommend reviewing quarterly, or whenever there is a significant change of content—for example, a new website design or new platform implementation (new web server or replacement content management system [CMS]). At the very least you should reassess your Google Analytics implementation and configuration on an annual basis. This review should coincide with your annual strategy and budget planning period, as it may affect those.

WHY CHANGE IS DIFFICULT, AND HOW TO OVERCOME RESISTANCE

Unfortunately, the most common outcome for any data project is stagnation. Once the initial excitement of collecting, organizing, and reporting on the data has dissolved, nothing else really happens. Yes, there are a few tweaks made here and there and maybe an organizational change (you hire someone to look after all this new data). You will have numerous data reports to review, but essentially the digital strategy remains the same as before. You may have guessed by now, that's an attitude I am passionate about changing.

I have yet to come across a senior executive who does not believe in the principle of measurement. But senior decision makers are not analysts. They have no real interest in the data itself—such as how it is collected, how it is cleansed, what advanced filters or segmentation options were applied, or what custom reports have been built. Those are not relevant to senior decision makers (they assume that has been taken care of).

On the other hand, what hypothesis you have, the conclusions and recommendations drawn from it, and an estimate of how reliable (accurate) those are, is of great importance to the business. In short, these are your *data insights*, the information and knowledge gained from studying the data that relate to the business. Executives crave this.

Insights are what executives require in order to make informed decisions and move the business forward. Unfortunately, this is where most analytics projects go wrong—they present data instead. By only having data to discuss at the decision-making table, you limit any impact. This is because without insights, the onus of understanding is placed on non-data experts. No disrespect to the knowledge and intellect of anyone in the executive suite (many are analytical thinkers), but interpreting data is not their role—how to use information is. Therefore, any data team has to make providing insights its priority. This means learning how to tell stories with data.

Analysts thrive on data; executives need insights.

The Importance of Storytelling

Communicating insights across an organization means using the data to describe a *story* that a person can directly relate to—in this case the senior decision makers of your organization. If you can communicate a compelling story with your data, you will get the executive buy-in to make the necessary changes for improvement, be it for your website, mobile apps, point-of-sales cross-sell performance, stock-keeping, event promotion—in fact, any Internet-connected device.[2]

An analytics team needs storytelling skills for two reasons:

- We are all resistant to change; one data point (or even several hundred) will not overcome this by itself.
- We all have a natural skepticism to data that does not show a direct causal effect we can relate to in our lives.

A clinical analysis of data can paint a bleak picture of underperformance. But on its own, data cannot take into account other indirect, though related, factors. It may be the case, for example, that performance is down in one area because the business focus was elsewhere at that time. Similarly, data can also give a misleading impression of success—for example, a competitor discontinuing a product at a time that coincides with your campaign launch for a similar product—in other words, luck. Hence data requires context.

Likewise, data without a context that you can relate to in your day-to-day job is just a number (the “so what?” effect). For example, the percentage of your visitors that add an item to their shopping cart—what does that number mean to you, me, or anyone else? It doesn’t tell us *why* the thousands of visitors who come to your website do, or do not, do this. Without an answer (a hypothesis), the recipient of data such as “38% of our visitors add at least one item to their shopping cart” is left with a “so what?” feeling. The data is meaningless to the business.

The obvious answer to the “so what?” question is the hypothesis that 38% of visitors are ready to buy, but in reality that hypothesis is often unfounded. People visit the site to check your delivery charges, stock levels, and the final price with all the additional charges that need to be added. In fact, checking the final price turns out to be very common user behavior for visitors of travel websites and for any visitor purchasing from an overseas supplier.

Building a story is about getting your analytical points across while also addressing both these issues. It is a skill that is acquired through experience rather than formal education. I provide storytelling examples in Chapter 10 (in ebook 2).

Using Data to Tell a Story

An expert proponent in the art and science of using data to tell a story is Hans Rosling.[3] He is a professor of global health at Sweden's Karolinska Institute. His work focuses on dispelling common myths about the developing world. Rosling's presentations are grounded in solid statistics illustrated by the visualization software he co-developed called Trendalyzer (www.gapminder.org). This innovative and free tool brings data to life by transforming boring statistics into animated and interactive bubble charts. It makes discussing global trends crystal clear and even fun.

Trendalyzer was purchased by Google in March 2007 (after the Google founders met Rosling at a TED talk). The tool is available as a Google visualization gadget[4] and incorporated into Google Analytics—both free and paid versions—and is known as the Motion Chart feature.

Hearing Hans Rosling speak is both educational and fun. I recommend the following TED talk as a taster:

www.ted.com/talks/hans_rosling_the_good_news_of_the_decade.html

THE WORK PROCESS

This section focuses on building a process for website analytics. However, it is equally applicable to tracking mobile apps or any other Internet-connected process.

Figure 3.2 shows how each of the four phases from Figure 3.1 flow into the next and the entire process is cyclical.

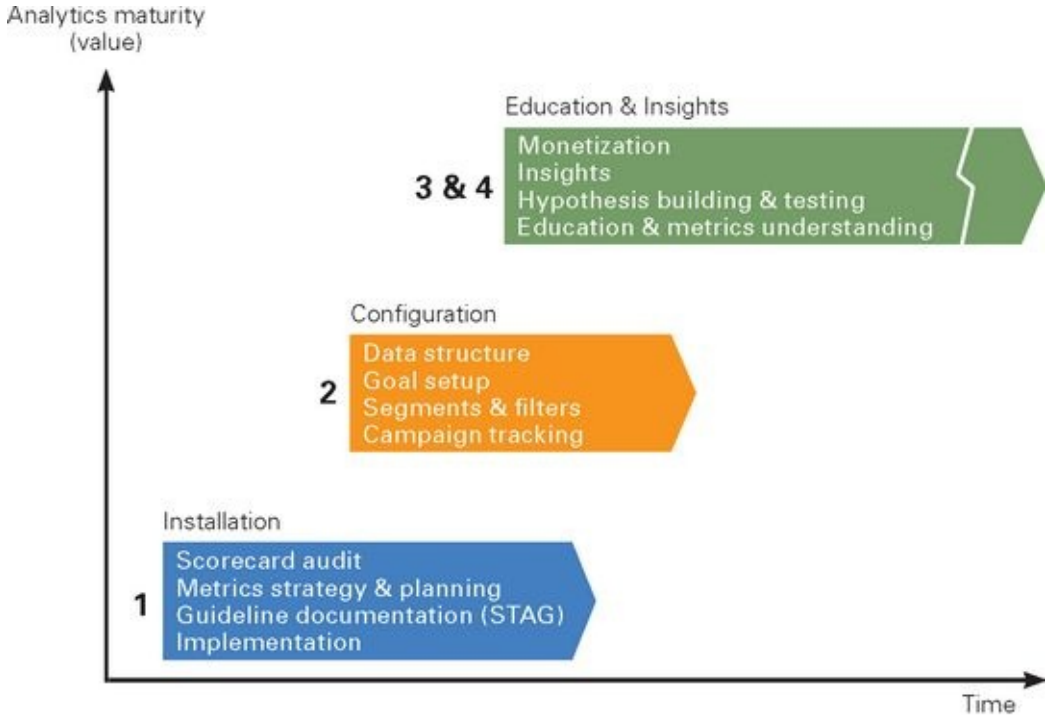


Figure 3.2 *The four-phase process in detail, with overlapping phases*

In addition to the breakdown of what work each phase requires, the flow in Figure 3.2 illustrates that although the order of the process is important, one phase does not have to be completed before the next can start, so you can maintain momentum for the project. For example, phases 3 and 4 run in parallel; as you gain insights you grow knowledge, and vice versa.

The metaphor I like when I describe this process is the building and launching of a ship (phases 1 and 2), and its initial voyages (phases 3 and 4). Both processes require a significant amount of planning and investment to get the project off the ground, involve multiple people to make it happen, and produce white elephants if they are not subsequently used by people to do real-world tasks.

Phase 1—Installation

The installation phase is where a great deal of your project's heavy lifting takes place (building the ship). Include all project *actors* and stakeholders in this phase. The actors include the analytics team and at least one member of the development team that will implement the resulting tracking code on your pages. Actors should also include a member of each of the departments relevant to the project that are represented in Figure 1.3—marketing, public relations, social media, sales, affiliate management, design, and content writers.

Phase 1 includes four steps:

- Scorecard audit
- Metrics strategy and planning
- Guideline documentation
- Implementation

Scorecard Audit

The scorecard audit is the first step. It assesses your current data situation—its completeness and its quality. If you have an implementation of Google Analytics already (getting the basic tracking code on your pages is straightforward), a scorecard audit is a 15-point checklist to highlight what you currently have set up and working correctly, as well as what is missing or not working. If you do not have Google Analytics in place, the audit is the same, though the content focuses on what needs to be tracked, rather than what is being tracked.

You can use the scorecard method to benchmark your data quality and to assess what is required for your organization to have a complete picture of its digital performance. Armed with this information, you will be able to ascertain how big a project you have taken on, and what resources, budget, and timeline are required to complete all the phases. In other words, you can set expectations. The scorecard is your blueprint for the project, and its importance cannot be overstated.

An example of a summary scorecard is shown in Figure 3.3. Note that once a Google Analytics account is opened and the standard tracking code is deployed across your pages, none of the remaining items (3–15) is tracked by default. You have to make adjustments to the tracking code, to your marketing campaigns, or to the HTML within your pages to make that happen.

Scorecard Summary				
		Weight	Status	Weighted Score
1	Account setup and governance	1.0		5
2	Tracking code deployment	1.0		10
3	AdWords data import	1.0		10
4	Site search tracking	1.0		0
5	File download tracking	1.0		0
6	Outbound link tracking	1.0		0
7	Form completion tracking	1.0		0
8	Video tracking	N/A	–	–
9	Error page tracking	0.5		0
10	Transaction tracking	2.0		10
11	Event tracking (non-pageviews)	1.0		0
12	Goal setup	1.0		0
13	Funnel setup	1.0		0
14	Visitor labeling	1.0		0
15	Campaign tracking	1.0		5
	Quality score (*QS*) out of 100			**27.6**

Figure 3.3 *Sample weighted scorecard to assess your data quality and set expectations*

Each of the listed items in Figure 3.3 is weighted according to its importance. That is, how important is it for tracking your visitor's user-journey? (How important it is for your business is discussed in the next section.) The visual presentation is shown as a traffic light system—red, yellow, green—to signify if there is a problem, or whether that item is reporting correctly. Obviously you will wish to see as much green as possible. The purpose is to show the status of each item but also combine these into a single weighted score for you to quickly assess your overall data quality. For the example shown, the quality score is 27.6 out of 100—that is, the implementation is only 27.6% complete. For this site, only the basics have been installed, and there is clearly a long way to go.

Producing a weighted scorecard audit is described in Chapter 4.

Showing Progress

In addition to benchmarking your current analytics position, the scorecard can be used as a data point itself—monitoring the progress of the first phase of the project. The scorecard is a useful illustration of what work is taking place behind the scenes in order to ensure that future analysis uses good quality data. This becomes increasingly important if your implementation timeline is long.

In some cases a best-practice implementation can take many months to complete (sometimes longer!). Having a regular score helps reassure senior executives that progress is being made, despite the fact that useful data is not yet available. When an implementation is protracted, I update the scorecard each month. If the score is not improving, an executive push may be required to ensure the necessary work is prioritized by your web development team.

Metrics Strategy and Planning

The purpose of the metrics strategy and planning step is twofold:

- For the analytics team to gain a good understanding of what metrics are required by the business
- To compare the metric requirements of the business with your current scorecard—what data is currently available and accurate

To achieve these goals, the analytics team sits down with the organization's digital stakeholders with the scorecard used to manage all parties' expectations. Before this stage, the weighting for each item on the scorecard is governed by how important the analytics team considers it for tracking your visitor's user-journey. Based on the discussions with your stakeholders, you may find yourself adjusting the weightings. For example, perhaps the downloadable brochure is more valuable to the business than you first thought? Getting agreement on the weighting (importance) of each row in the scorecard of Figure 3.3 is an important achievement in this step.

Keep in mind that this is early days in your project. The purpose here is not to set a metrics strategy in stone. Rather, it is a plan to point the flexible analytics ship in the right direction and manage the expectations of what is required to achieve this. Improvements can be added at a later stage (and should be). What is being set up in this step are the solid data foundations that will support the future modifications and additions.

With the scorecard agreed to and expectations understood by all stakeholders, you can now start to plan your implementation—that is, who does what and when to make it all happen.

Guideline Documentation—the STAG

Documentation is often neglected, but it's critically important that your installation be documented. By that, I do not mean detailing the code or the precise page and server logic required to set up your tracking. That would be a tedious read and would be out of date quickly. Rather, you should describe the reasoning for the current setup and the proposed data structure in a site tracking assessment and guidelines (STAG) document. This is the document that will go to your developers to implement the Google Analytics code.

Who Writes the STAG Document?

The STAG document is a technical document and hence a person with Google Analytics developer experience is required. This could be a member of your existing development team, though most likely you will employ the services of a GACP. The role of a GACP is discussed in Chapter 2, specifically with reference to Figure 2.3.

Why is a STAG document necessary?

In Google Analytics there are at least three different tracking methods available to track the same thing—pageviews, virtual pageviews, and events (for more detail on these see the appendix and Chapter 6, in ebook 2). Which of these methods is best for your implementation will depend on many factors, most of them technical—for example, your CMS setup, web platform architecture, content structure (for example, embedded third-party content), and the complexity associated with changing any of these. In addition, each tracking method has its own data structure. Hence it is important to document which methodology is employed for each of your tracking requirements—each item in your scorecard.

For example, if each scorecard item were to be considered in isolation, a different tracking method might be proposed for one that made it difficult to compare with the others. That is a nightmare scenario that leads to comparing apples with pears, with bananas. By having a central document that outlines the tracking choices made, the reasoning behind the choices, and the expected data structure, you greatly simplify any future tracking developments, troubleshooting, and analysis. The STAG document ensures all those involved in your analytics project understand the soundness of what data is being collected, *without* having to read code on web pages.

The STAG—the Analyst's Reference Document

A web analyst's first question when investigating an unexpected change in traffic patterns is "How was this data collected?" Analysts need to understand whether the change is real or perhaps an unforeseen consequence of a page edit that inadvertently altered the tracking. Without first answering this fundamental question, a great deal of time and effort can be wasted going around in circles. The STAG document is therefore the first port of call for the analyst in order to understand exactly what may be causing the observed change.

A STAG document should contain approximately one page of explanation for each item in the scorecard—a total of 15 pages. Each section should be written in a self-contained manner to help the reader refer to a specific section without the need to read the entire document as a novel. It should also include screenshots and diagrams where possible, as shown in Figure 3.4.

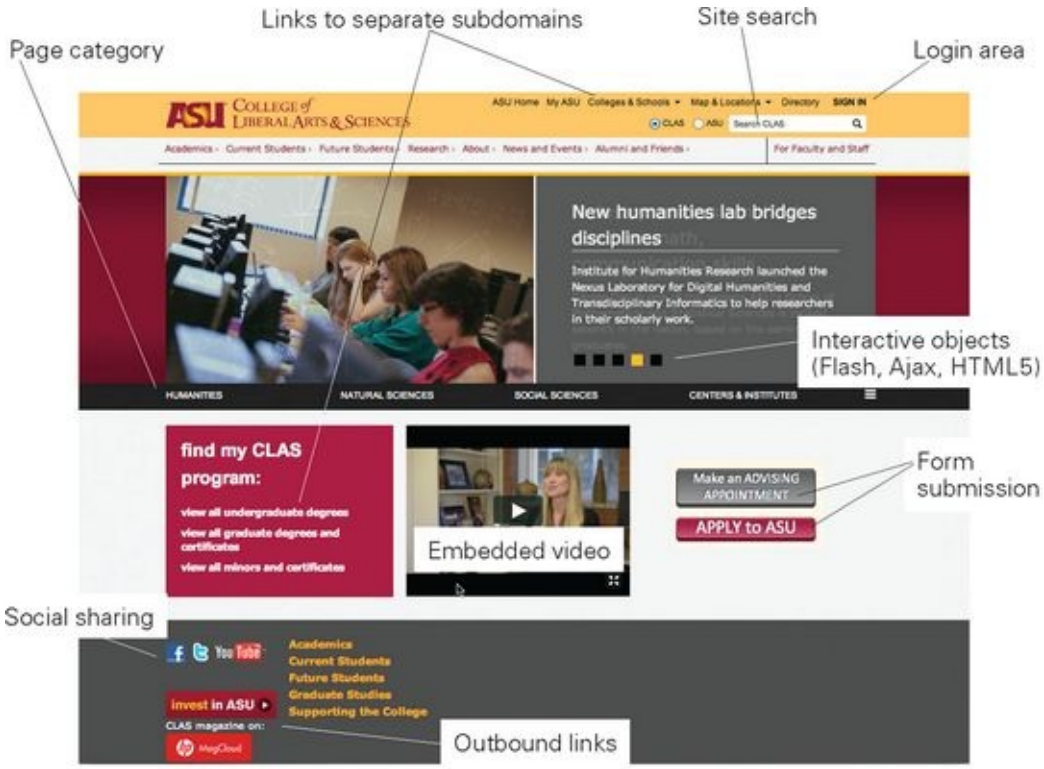


Figure 3.4 *Sample webpage screenshot highlighting the features that require special tracking consideration*

Although code examples may be necessary within the STAG—this document is after all going to your developers—the code should be limited to the standardized tracking code snippets that can be found at http://developers.google.com/analytics/devguides/collection/.

Figure 3.5 illustrates the concise nature of a single STAG item—in this example, for the requirement to track file downloads of different types (PDF, DOC, ZIP files). A few things to note from this example:

5. File Download Tracking

For ALL file downloads e.g. PDF files as listed at http://www.site.com/product/shortthrow/vpl-bc825/overview, modify the onClick event handler so that an **event** is sent to Google Analytics as follows:

```
ga('send', 'event', 'Downloads', 'PDF', '$fileTitle');
```

Where `$fileTitle` is the PDF file title.

For example:

```
<a href="/attachment/1237486044794" onclick="ga('send', 'event',
'Downloads', 'PDF', 'VPL-SW125 Datasheet');"
name="ATTEnEU201205FPJShortThrowVPLbc825" class="button-type1" title="VPL-
bc825 Datasheet"></a>
```

In this case, `$fileTitle` = "VPL-bc825 Datasheet" (not the file name1237486044794)

If there are other file download types e.g. doc, xls etc., replace the string "PDF" in the code accordingly.

A CMS technique to simplify the tracking of all file downloads (rather than finding these manually), is to define a separate HTML class attribute for your file download links. Then apply the onClick event handler to all links matching this class. Currently your PDF files have the class attribute "`pdf-items box-type-1`", so this may be straightforward to implement.

Figure 3.5 *Sample STAG section specifying the requirements for file download tracking*

- The entire description of what is required is covered within half a page.
- The necessary code changes are highlighted in bold (non-bold code is the

existing page HTML).

- Having outlined the data structure (using an event), I leave it to the developer to figure out how to capture the file title. The filename is meaningless in this case and cannot be changed.
- The example is for tracking PDF file downloads. I leave it to the developer to apply the same technique to other file types.
- The standardized code is shown as being manually placed within the HTML link. However, a developer will be familiar with other *listener* techniques that can automate this—for example, using a tag management system such as Google Tag Manager.[5]

The STAG item shown in Figure 3.5 conveys the data requirements and data structure clearly and concisely, while allowing the developer the freedom to figure out the best way to achieve the outcome. Ultimately, it is up to your developers to decide on the best method and what the exact code should look like (that's what you pay them for).

Implementation

The implementation of tracking code is the responsibility of your development team—either internally or via your agency. It is the final step of this phase, and it should result in good, clean data being populated into your Google Analytics account. Assuming a member of the web development team has been involved throughout this phase, they should be expecting to receive the STAG document.

It is important that whoever implements the STAG realizes this is *not* a cut-and-paste operation. The developer is required to think, and time needs to be allowed for this—that is, to understand the data requirements, the thinking behind them, and the standardized code snippets presented. They then build their own version of code to go on your pages. How to deploy the code and the logic required to obtain the necessary parameters are your developer's responsibility. See the sidebar "Using a Tag Management System."

As any person familiar with a Google Analytics implementation can tell you, the required tracking code snippets are straightforward for any developer. After all, millions of blogs, hobby sites, and mom-and-pop businesses use Google Analytics around the world and certainly did not go through the detailed process I am describing in this chapter. However, at an advanced level—beyond items 1 and 2 on your scorecard—a more detailed implementation approach is required.

Using a Tag Management System (TMS)

Think of a TMS as functioning in the same way your CMS works—except it is for managing snippets of code (tags) rather than content.

A TMS, such as Google Tag Manager,[5] works by being deployed on your pages as a *container* script. This is a JavaScript snippet—typically a dozen or so lines long. Once in place, the container allows you to insert tags (other code) into your pages via a centralized, rules-based system. The benefit is that you

remove all other snippets of code and replace it with one that is edited via a separate web interface. Any additions or modifications to the tags do not require changes to your pages. Nirvana for your IT team!

If you use a TMS for your Google Analytics implementation (I strongly recommend you do), then your IT department's role can be as straightforward as deploying the TMS container on all of your pages—a one-time deployment. The analytics team then implements the STAG via the TMS web interface.

Phase 2—Configuration

While phase 1 (installation) required the inclusion of all actors, phase 2 is covered by the analytics team working in conjunction with the marketing and communication teams. In phase 2 you are adjusting and fine-tuning your Google Analytics implementation. Using my shipbuilding analogy, this phase is the outfitting of the decks. The necessary work is completed either within the Google Analytics interface or during the creation of your marketing campaigns. No modification of your web pages is required (your web development team can take a break).

Phase 2 includes these four steps:

- Data structuring
- Goal setup
- Segmentation and filtering
- Campaign tracking

Data Structure

Data structure affects enterprises with multiple brand-specific or product-specific websites, or a website targeting different markets. If this describes your organization, then it makes sense to have separate, standalone Google Analytics reports for each website or market. That way, a product manager or country manager can focus on the reports relevant to their product or country. By *standalone* I mean that your product or country managers do not wish to see data from other products or markets mixed in with their reports. For example, they will not want reports where euros are mixed with dollars, or reports including visits from different time zones if they are not relevant, as these will skew the data. By keeping reports separate for these scenarios, analysis becomes easier.

On the other hand, enterprise users also have a headquarters that requires a high-level overview of how the *entire* online channel is performing. This is when having separate report sets is laborious—you literally have to open multiple browser windows to view the same reports for each product or market. To avoid this, set up a *roll-up* report. That is, in addition to individual market or product reports, you have a single catch-all report set with data from all websites or markets aggregated. You achieve this by adding a second Google Analytics tracking code snippet to your pages. In this way, all data hits are sent to two reports sets (Figure 3.6).

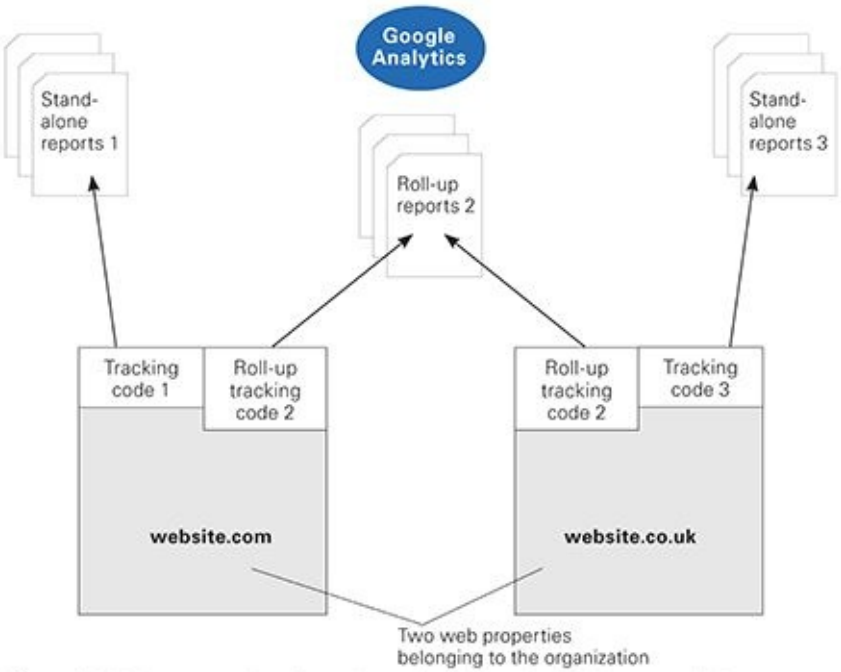


Figure 3.6 *Using a second tracking code to simultaneously send data to multiple report sets*

Using the multi-tracker setup shown schematically in Figure 3.6, the two websites have their own data reports (report sets 1 and 3), while also sending data to a common roll-up report (set 2). The principle is straightforward, though careful consideration is required if combining currencies and media spend. For example, you will need to convert all currencies by a fixed exchange rate so that the monetary totals make sense. The exchange rate does not have to be exact, but it is important it remain fixed. Otherwise your sales performance will vary based on exchange rates and not on any improvement in your content or marketing.[6]

Which approach you take to structure your Google Analytics data depends on how your digital content is structured. It may be that your data is already aggregated and you need to figure out how to separate the data while maintaining the integrity of the visitor journey (see sidebar “Visitor Journey Integrity”). Alternatively, maybe your data is already split and you require an aggregation method.

Chapter 7 (in ebook 2) discusses geographic data structuring in more detail.

Visitor Journey Integrity

The common technique for separating data is to use *view filters*. These manipulate data at the hit level, not the visitor or session level. Care is required when manipulating data at the hit level. For example, the first data hit defines the start of a visitor’s session and contains all the referrer information (where the visitor came from). If an applied filter removes the first data hit, you will observe odd data in your reports—such as visitors having no referrer data, or visitor sessions being restarted (duplicate counting).

Goal Setup

Goals are the important engagements happening on your website that distinguish

a genuinely interested visitor from a tire kicker. A goal is a call to action—what you want a visitor to do when they see a particular page or content within the page. An obvious goal for a transactional site is a purchase, and this should be set up so you can analyze the path on your website that leads to the sale (the sales funnel). However, this goal is black and white—visitors either purchase or not. The number of visitors who purchase from you is usually a small fraction of your total visitor volume (typically around 3% of your traffic[7,8]). So while the purchase goal is important, it is a small piece of your website's potential.

My point is that your goals become most informative when used to understand visitor interest—those engagements that help you *build* a relationship with what otherwise are anonymous visitors. Regardless of whether you have an e-commerce facility, use goals to measure visitor engagement with your brand and products. Examples of these engagements are listed in Table 3.1.

Table 3.1 *Goal Examples*

Goal Type	Visitor's Action Defining the Goal
E-commerce	• Purchase confirmation • Add-to-cart action (e-commerce site) • Got to step 3 of 4 in the purchase funnel (came very close to becoming a customer) • Transaction failed—a negative goal (one you would rather see less of)
Lead generation	Any action where personally identifiable information is passed on: • Completed a contact request form • Clicked a mailto link • Subscribed to your newsletter
Purchase intent	• Logged in to your site • Visited your store finder page • Clicked an outbound link to a reseller
Brand engagement	• Logged in to your site • Downloaded a file—such as your product brochure or price list • Viewed a specific page, such as a special offer • Watched a video clip • Watched a video clip to completion (or passed a threshold of *x*%) • Shared or commented on your content on Facebook, Twitter, Google+, LinkedIn, your blog, and so forth (social sharing) • Used a widget (such as a loan calculator) • Viewed *n* or more product pages during a long visit (where you have set a threshold, such as $n > 10$ and time > 5 minutes) • Used advanced features of your internal site search facility (examples: more than one related search; using filters) • Used any content feedback or rating mechanism
Other revenue	• Clicked on a third-party advertisement

Think of goal conversions as specific measurable actions that can be applied to every visit. Up to 20 goals can be configured within Google Analytics, and your target should be that at least one goal is completed by a genuinely interested visitor. Otherwise it is fair to ask the fundamental questions: What are these visitors doing here? Is our traffic acquisition sending us poorly qualified visitors? Is our website content good enough? Is the user experience poor?

The value to your organization of each goal completion will vary. Some goals

may indicate only a small interest and hence have a low value—a PDF file download, or pressing play on a demo video, for example. Others will be stronger indicators—such as watching a demo video to completion (or viewing 80% of it). Therefore, not all goals are equal.

To differentiate your goals, monetize them. That is, apply a monetary value to each goal completion. Monetizing your goals is discussed in the section on phases 3 and 4 later in this chapter.

Defining Success—Macro and Micro Goals

Defining website goals is probably the single most important configuration step you will apply in Google Analytics, because it is how you define success. Goals are typically the reasons why your website was built in the first place. For example, was it to sell directly, to generate leads, to keep your clients or shareholders up to date, to provide centralized product updates, or to attract new staff? As you begin this exercise, you will realize that you actually have many website goals. These are your *macro* goals.

Goals don't have to include the full conversion of a visitor into a customer—that is obviously very important, but it's only part of the picture. If your only goal is to gain customers, then how will you know just how close noncustomers came to converting? You gain insight into this by using *micro* goals. These measure the building of relationships with your visitors.

Goals are the building blocks of key performance indicators (KPIs), which are the key metrics your organization focuses on to assess digital performance. How to establish KPIs is covered in Chapter 9 (in ebook 2).

Segmentation and Filters

Segmentation and filters are the tools you use to group together visitors of similar behavior (segments) or cleanse your data (filters). For example, a segment of visitors could be your *customers*, *high-value customers*, *subscribers*, *noncustomers*, or *engagers*. I use the term *engagers* to describe visitors who have done something on your site that indicates significant interest in you but who are not yet customers. Downloading your brochure can signify engagement, as can filling in a contact request form, commenting on a blog article, or socially sharing your content. You can even combine these criteria into a sequence segment—visitors who match criterion *X* and immediately match criterion *Y*.

Segmentation is the technique employed to focus your analysis on groups of visitors who have similar characteristics, rather than on the whole data pot. The purpose is to improve the signal-to-noise ratio by bubbling up data points that otherwise lie below the surface (and potentially go unnoticed). Understanding how a handful of visitor segments behave is far more practical than trying to comprehend the thousands of individual visitor patterns that exist on your site. On the Internet, with so much variability, almost everyone exhibits unique behavior. Understanding segmentation and its importance is discussed in Chapter 5.

Filters can also be used to group and segment data. However, their common

use is to cleanse data. Perhaps your web page can be displayed as /Products.htm or /products.htm. To Google Analytics these are two different pages and are therefore reported separately, which is not helpful to anyone viewing the reports. By using a lowercase filter, these can be combined into a single page with all page metrics aggregated. Similarly, you may wish to remove meaningless parameters from your page URLs (such as session IDs). By removing these you make your reports easier to read and interpret.

Other things you can do with *view filters*:

- Remove visits from your own staff or agencies
- Rename obscure pages into something more memorable
- Separate visitors geographically so that you can, for example, view reports from US visitors and visitors from other countries separately
- Ensure campaign tracking is applied consistently (see next section)
- Rewrite legacy campaign tracking parameters for Google Analytics (rather than having to reconfigure existing campaign links)

Campaign Tracking

Campaign tracking is the technique to measure the performance of your digital marketing efforts—from *all* channels. It answers questions like these: What drives visitors to our website and mobile apps? Which are the most effective drivers? How do different campaigns compare against each other? Typically your digital marketing costs are much greater than the cost to develop, build, and maintain your digital presence. Knowing the performance of your marketing and determining which campaigns produce the best results are key advantages of web analytics. You probably know that half of your marketing works. But which half? In the digital world, all marketing can be tracked. But campaign tracking does not happen by default. You therefore need a rigorous system in place to ensure it does.

I discuss the setup of campaign tracking in Chapter 6 (in ebook 2).

Phases 3 and 4—Insights and Education

On reaching phase 3 you will be on the brink of crossing the dotted line shown in Figure 3.1. That is, your analytics project is moving from being a cost to the organization to one that generates a return. The extent of the return can be quite extraordinary. For example, according to a compilation of statistics by the Baymard Institute,[9] the average *shopping cart abandonment rate* is 67.44%. That is, of all the online visitors that have added an item to their cart, two-thirds leave without making a purchase.

Similarly, average *checkout abandonment rates* are reported at 43%.[8] That is a staggering number—43% of visitors that have completed their checkout process and are ready to pay (have their money in their hands, so to speak) bail out of the decision at the very last step. Clearly, with such significant sums of money being left on the table, any insights that can help improve conversion will have a significant dollar impact on the business. This was the basis of the work shown in Figure 1.2. In that example, the improvement resulted in a fourfold increase in revenue for the website, equivalent to $7.5 million extra per year (a significant amount for the boutique travel company concerned).

What about insights for non-transactional websites? This is where monetization comes in.

Cart Abandonment versus Checkout Abandonment

The *cart abandonment rate* is defined as

$$\frac{\textit{number of purchases not completed}}{\textit{number of purchase initiations}}$$

Often visitors will add an item to their shopping cart in order to check your final price, or even to obtain a price (this is common in the travel industry). Hence, the cart abandonment rate can appear to over-inflate the abandonment problem. That is, some visitors will leave the shopping area of your website for reasons not connected with their user experience.

The *checkout abandonment rate* is defined as

$$\frac{\textit{number of purchases not completed}}{\textit{number of checkout initiations}}$$

The checkout abandonment rate measures abandonment from when people are asked to make their payment—it is the penultimate step in the sales funnel and is not affected by visitors who check prices, stock levels, or delivery options. A high checkout abandonment rate is a clear indication that your visitors

are experiencing problems—either technical issues (error messages or some kind of failure); a lack of confidence or trust in your ability to fulfill their order; or other reasons, such as a high shipping price or long delivery time that they were not expecting from the start of the sales process.

Monetization

The purpose of monetization is to ascertain the *value* of your website and keep track of it over time. I cannot overemphasize the importance of this step. Whether you have an e-commerce facility or not, your website has value, as defined by

total website value = e-commerce revenue + goal value

E-commerce revenue is tracked by the addition of specific code on your purchase confirmation page. It is the responsibility of your web development team and is a part of phase 1. On the other hand, your goals are the calls to action that the analytics team defines in phase 2 (see the section “Goal Setup,” earlier in this chapter). The definition of goals takes place within the Administration area of Google Analytics. No changes to your web pages are required. In the monetization step we consider adding monetary values to your goals.

As discussed in the “Goal Setup” section (refer also to Table 3.1), the value of each goal completion will vary for your organization. Some goals—for example, the viewing of a special offers page—have zero value. They may be important for your new visitor acquisition strategy but have no tangible value. Other goals have a high value—for example, any goal that results in the visitor sending you their personal information (such as their name or email address), as this allows your sales team to contact them and work their magic. And there are goals that fall in between. Thus, not all goals are equal; you account for this by setting goal values.

Goal values are specific to your organization and therefore are not arbitrary amounts—they require some thought, though the process is straightforward. The process is the same for all your defined goals. Evaluate how often visitors who complete a certain goal go on to become customers and what that customer is worth. For example, if your sales team can close 10% of people who submit your contact request form and your average transaction is $500, you would assign $50 (10% of $500) as the goal value to your contact form request. In contrast, if only 1% of submissions result in a sale, you assign $5 to your contact form request.

Your goal values should represent a real monetary amount, though it will be approximate. Real monetary amounts are obviously preferred, so that your organization can directly relate to them. However, they can be difficult to ascertain. If this is the case for your organization, use a relative scale where the values are approximate, yet still relatable. For example, watching a demo video to completion is worth 10 times the value of a PDF download. Table 3.2 shows

examples of common goal values.

Table 3.2 *Using Relative Goal Values*

Goal Type	Example	Relative Value
Zero value (important touch points with your visitors, but have no tangible value to the business)	• Viewed a special offers page or other specific URL • Watched a video clip (no threshold) • Used a widget • Viewed *n* product pages over a period of *m* minutes • Used your internal site search facility and did not receive zero results	0
Low value (a strong signal of visitor engagement, though nothing to indicate sentiment)	• Downloaded a document (PDF brochure, specification sheet) • Watched a video clip to *x*% completion • Clicked on a third-party ad • Used any anonymous content feedback or rating system	1
Medium value (a strong signal your visitors like or value your content and are likely to purchase)	• Shared socially (Like, Tweet, +1, Follow, Pin, and so forth) • Visited your store finder page • Clicked an outbound link to a reseller	5
High value (the ultimate engagement, allowing you to connect and talk directly with your visitors)	Provided any personal contact information, such as • Form completion (contact request form) • Newsletter subscription • Clicking a mailto link • Adding a product review • Adding a blog comment	10

By monetizing the goals on your website, you achieve two important things using Google Analytics:

- You can readily report and monitor the total value of your website over time.
- Google Analytics automatically calculates and assigns a value to *every* page viewed and *every* visitor you receive.

That latter point is a great asset to focus on for analysis. It is how you identify your high-value visitors, high-value pages, and poor-performing pages. Understanding value as reported in Google Analytics is discussed in Chapter 6 (in ebook 2).

Insights

Google Analytics is your main tool to help you understand, and therefore optimize, how visitors arrive at your website and what happens once they are there. This is what Google Analytics was designed for.

Providing insights is about providing the reasons for change. The analytics team is therefore the agent for change in your organization—change in how the digital marketing is conducted (for a better return on investment) and change in the website content or its architecture (for a better visitor experience). Getting these right determines if your website takes off successfully or remains lackluster.

Reasons for change need to be based on solid thinking to be credible—that is, analyzing good quality data and building sensible hypotheses from it (or using the data to confirm an existing hypothesis). Data insights come is two flavors:

- Deductive insights—having an existing hypothesis, you search for data to support it.
- Inductive insights—you argue backwards from a set of observations to a reasonable hypothesis.

Both approaches are valid, though I have found from experience that inductive insights are more common (I estimate 2:1).

Insights enable you to understand how to attract highly qualified leads (high-value visitors) to your website and what they do once they land on your site (convert or not). It is a powerful skill set to develop. Strongly overlapping—though separate—industries have developed around them—digital marketing and conversion rate optimization, respectively. My view is that these have separated because the sorts of people they attract—marketers on the one hand, technical people who manipulate website content and its architecture on the other hand—don't overlap much. In my view a good digital analyst needs both skill sets (see Chapter 8, "Building Your Insights Team," in ebook 2).

Hypothesis Building and Testing

Data is only useful if there is a hypothesis to describe what it means to the business. Otherwise, it's just noise. The converse is also true— having a hypothesis requires supporting data. You achieve this by designing experiments (tests).

For example, a hypothesis could be that your product photograph is the most important driver for a visitor on your website to convert into a customer—more important than the campaign headline or surrounding text. You can test this (provide data to support or refute the hypothesis) by showing half your visitors the original page layout and half with the image removed. At the end of the experiment you compare the conversion rates for each visitor segment. Did visitors who saw the page with the product image convert at a higher rate than those who did not? This is the basic design of A/B testing.

Taking this a step further, if the image is found to be a key component of the conversion process, you could test the effectiveness of different images. You might have 20% of your visitors see each of five alternative images. If the test data shows a significantly higher conversion rate when image 2 is shown, you know that only image 2 needs to be used on your web page for this product. You have just scientifically improved the performance of your website based on data.

Often, the conversion process is more subtle than relying on one image. Maybe there are multiple images required to tell a story to your visitors, or the product image, headline text, and surrounding descriptive text are all important and work

together to procure the conversion. In this case, different combinations can be tested together. For example, you could test three product images, three headlines, and three descriptive texts (27 combinations in total). This is the basis of multivariate testing.

Testing with Google Analytics There are two types of testing to consider:

- Offsite testing—experimental tests you perform with your marketing campaigns to attract visitors, such as testing different banner designs or email newsletter layouts. The experiment is conducted via a third-party vendor tool, though it is straightforward to label the visitor once they click through to your website. Google Analytics can report which test version they clicked on, and you can know which version is performing better once the visitor is on your website.
- Onsite testing—experimental tests you perform with visitors on your site, such as testing the effectiveness of different product images, descriptions, prices, and page layouts. Google Analytics includes a testing platform called *content experiments* that can conduct A/B tests. However, it is also possible to use third-party tools and integrate the results *within* Google Analytics so that reports show which version is performing better.

The Google Analytics Content Experiments platform allows you to perform simple A/B, A/B/C, etc., tests on your webpage content. It is a solid platform that can be deployed with an interesting statistical method—the multi-armed bandit model. This optional setting helps you obtain results faster than more traditional statistical methods.[10] Reducing experiment time is a key advantage of content experiments. However, in my view, a drawback of the tool is the way experiments need to be implemented on your pages. For all but simple experiments they require your web development team to reengineer how your pages are delivered. That is, versions are delivered server-side.

I find the server-side approach of content experiments is not a viable proposition for enterprise-level web development teams—it is too big of an overhead and adds a layer of complexity to running the website. As an alternative, specialist testing platforms are available that use client-side methodologies. This simplifies deployment while still integrating test results within Google Analytics.[11]

☛ *I discuss my reasoning for avoiding Google Analytics Content Experiments at* http://brianclifton.com/avoid-content-experiments

When not to test For data to be considered *useful* it must have a hypothesis to describe it. However, not all hypotheses require data to support them. That may sound odd, but sometimes test data is too difficult to obtain and performing an experiment is not worth the effort. For this case, expert judgment (that is, experience) can trump a lack of data. Not everything that can be tested should be tested.

If you were to test every possible combination of content elements on your site, you would be waiting beyond your lifetime for the results. The quantity of data would be overwhelming. As with all things digital, the medium moves fast. Hence, rapid incremental testing is far more effective than long-term testing (an approach that is embedded in Google's product philosophy).

Here's a real example from my days when I was working at Google, when using good judgment (experience) should have been chosen instead of hypothesis testing:

I was looking to hire. My manager in his infinite wisdom (now a senior executive who shall remain nameless) asked me why he should invest in my team. "Google Analytics is being given away for free. What more could a user want?" I explained that with more resources I could go beyond just giving away the product and actually help users make sense of the reports and so actually benefit from the product. Put simply, I wanted to help users better understand Google Analytics and hence their websites. I hypothesized that if they did, they would go on to invest more in their digital marketing. As Google dominates digital marketing, this would ultimately drive more revenue to the business, right?

"Prove it. Prove that there is a link between increased digital advertising spend and Google Analytics usage."

I spent the next two years attempting to show a correlation by testing—comparing the digital ad spend of clients with and without Google Analytics. Intuitively most people would agree with my hypothesis. So far I have not met anyone who disagrees. However, it was incredibly difficult to show any correlation. At that time, circa 2006, all things digital were growing at an exponential rate. Take into account seasonality, new advertising features from Google, changes in the competitor landscape, other tracking technologies, the growth of smartphones, social media, and more; and I was never going to succeed in building a solid data foundation for my hypothesis. The issues were too complex and the causality too easy to doubt.

My point with this story is to emphasize that testing different alternatives, just because it's possible to do so, does not mean it is a worthwhile exercise. In fact, it can be costly to your organization—I spent two years going around in circles and missing other important opportunities, such as gaining support from other stakeholders who could have invested in my team.

When to test To decide if testing is required, first conduct your analysis and build your hypothesis (or vice versa) about the visitor behavior on your website. This is the most important step of any experimental setup. Your default position should be to test only if and when it is required.

Before considering if a test is required, ensure that best-practice user experience principles are applied to your website's content and structure. Do not skip this process. Conducting a test experiment can take weeks, and sometimes

months. Fixing user experience issues is usually quicker, easier to scale (improvements can be applied to all your pages via templates), and negates the need to test. For example, you do not need to test what is obviously broken.

In other words, experiment by testing the important questions that matter to your business. Here are some examples.

Testing for marketing (offsite testing)

- Different banner designs—for example, static versus animated
- Different email newsletter layouts and designs
- Different call-to-action texts

Testing for design and persuasion (onsite testing)

- Different value proposition statements or special offers
- Different trust factors, such as logos, certifications
- Different call-to-action texts

Testing for usability (onsite testing)

- Revised content layout, such as shorter forms, compacted checkout process, reduced scrolling
- Simplified navigation, such as fewer clicks to destination, better site search results
- Comparing text links versus buttons versus images

Fix the User Experience First

Web developers make things work, web designers provide the creative flair, and digital marketers bring in the traffic. However, it never ceases to amaze me how little attention is given to the user experience.

For example, it's not always obvious that call-to-action buttons are just that, a call to action. Often such buttons are graphics. From a visitor's perspective, these appear to be a part of the design aesthetic. They do not look like a clickable button, and therefore suffer from *banner blindness*. The buttons are ignored—considered ads by your visitors, rather than important calls to action.

In addition, the result of clicking the button—what happens next—should be completely obvious to the visitor. The principle is called "Don't make me think." This approach is explored in the book of the same name by Steve Krug.[12]

Education and Metrics Understanding

From the start of phase 3, I recommend education running in parallel with your project. The more you democratize your data—that is, make it available throughout your organization—the more buy-in you will gain for your recommended changes. My view is that data should never be treated as a silo—it is too valuable to the organization. And having the wisdom of the crowd is an

asset to be appreciated.

However, with data democratization comes an overhead—questions: What does this number mean? How can I use this number? Can you pull the same number but segmented by *y*? These questions will come through, first as a trickle, later as a torrent. After all, information is empowerment—people make better decisions with more information. So wanting more is a natural human behavior.

Create power users Rather than drowning your team in the work of answering all questions, educate your interested parties so they can self-serve. Raise their level of metrics understanding so that they can view the reports in Google Analytics themselves and answer their own questions. For example, in an organization with multiple regional or country-specific offices, train one suitable person from each to serve the data needs of that office. That person should be involved with the digital marketing team and be numerate and digitally savvy. They become the bridge between the data-intensive analytics team that is focused on the central digital strategy and the local office that has different day-to-day needs. These are your Google Analytics *power users*.

Power users are your data advocates. Once trained, they know the Google Analytics product well enough to self-serve, which means they absorb the more basic reporting questions on your behalf. They will understand the direction of the analytics team, yet they are embedded within their own business unit. That's a great combination—it means your power users can guide data expectations locally and escalate to the central team when appropriate. See Figure 3.7.

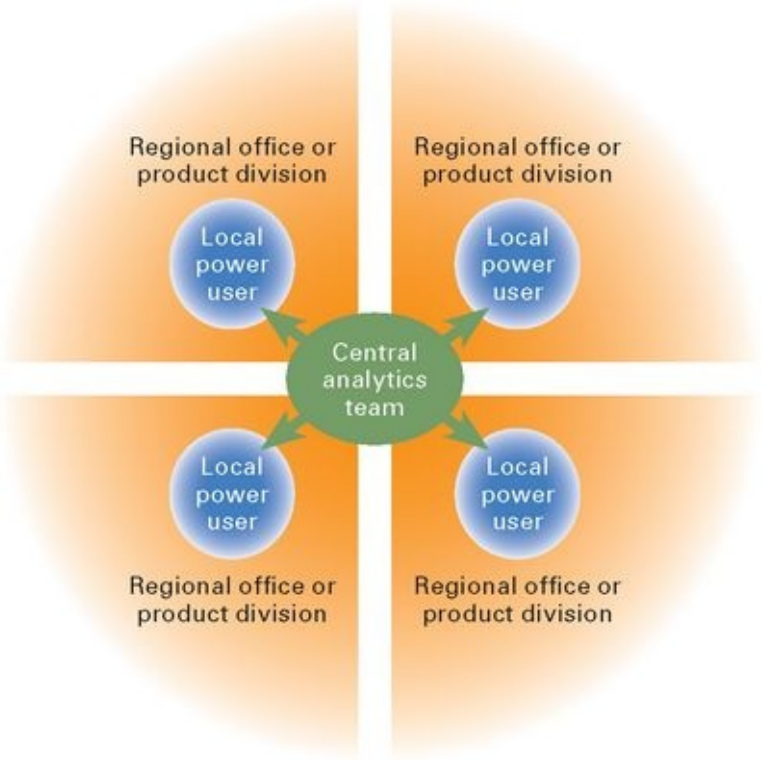


Figure 3.7 *Democratizing the data using Google Analytics power users*

Keep in regular contact with your power users—I recommend a conference call for general updates and Q&A every two weeks. Provide continuous formal training to enhance their skill set—preferably once per year. With an engaged and

effective power user network in place, you will probably find they are your next team hires. Good analytics people are not easy to find.

Building an analytics team and utilizing Google Analytics power users is discussed further in Chapter 8 (in ebook 2).

HOW LONG WILL THIS TAKE?

How long will it take to move from ground zero to the stage where insights are produced? In Figure 3.2, a realistic time frame is to allow one quarter for each step. That is, by Q3 you should be in a position to start gaining real insights for your organization. Of course, that is an estimate that can vary greatly. However, I recently looked at over 50 client engagements, and this time frame remains a good estimate.

Factors that alter your timeline Of the four phases shown in Figure 3.2, the one that requires the most heavy lifting is phase 1— installation. Once this is completed, the remaining phases can go quickly. A basic installation—items 1 and 2 in the scorecard shown in Figure 3.3—is something that can be achieved in a matter of hours for a good web developer. It can be even less if what you have is a single website domain without an e-commerce facility. However, tracking multiple websites, subdomains, e-commerce facilities, mobile apps, non-web devices (checkout tills, barcode scanners), and all the other items listed on the scorecard requires a good deal of thought and planning in order to get the implementation right.

Beyond these, timing depends on the resources and priority you can allocate to the project, the platform your content is hosted on (some CMS systems can be painful to work with if they have not been set up with tracking in mind), and who performs the implementation. The time dependence of the last variable (who) is surprising. I initially expected organizations with an internal web development team would respond the quickest. However, it turns out that organizations that employ a third-party web development agency are more agile—that is, faster to implement—on average.

Why would an outsourced third party be faster at implementation? For a third-party web development agency, there is a finite time they can allow for the project's completion—the longer it takes, the less profit they make. On the other hand, internal web development teams tend to suffer from a number of issues that result in the analytics implementation taking longer—internal politics, lack of tracking prioritization, lack of interest (tracking is not what they do or want to do), and lack of ownership. They view tracking as a function of marketing.

Bring the web development team on board as a key stakeholder from the beginning of the project. Their role does not end when phase 1 is completed. They are needed to implement tests as well as any resulting changes you require. It is a long-term relationship that requires good communication between both parties.

BUDGET—WHAT DOES IT COST?

How much should you budget for web analytics? This question can be interpreted in two ways:

- What do other users spend on their web analytics?
- What should *we* invest in web analytics?

To answer the first question, review Figure 3.8, taken from Econsul tancy's "Online Measurement and Strategy Report 2013."[13] The amount spent on web analytics covers the costs for internal staffing, technology (tools and implementation), and the use of third-party consulting services. See Figure 3.9 for the expenditure split.

Figure 3.8 *How much do users spend on web analytics each year?*

Source: Ecoconsultancy

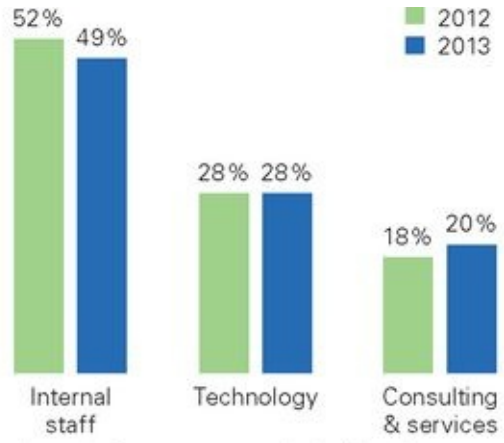


Figure 3.9 *The areas web analytics users spend their budget on*

Source: Ecoconsultancy

Figure 3.8 shows that more than half of the survey respondents spent less than a total of £10,000 per year (approximately $15,000) on their web analytics. With the free version of Google Analytics, this is expected. My reasoning is that users consider a free tool as not worthy of much investment. Conversely, if the exact

same tool has a high license cost, I predict that more expenditure will be committed to ensure it is used effectively—psychologically, no manager wants to have their purchase viewed as a white elephant.

To answer the second question, if you take your data seriously then you need to invest in your web analytics. The free version of Google Analytics is very powerful and suitable for most organizations. However, you still require internal resources (trained staff), technology (ensuring a continuous supply of good quality data), and consulting services (third-party expertise) if you wish to make informed, strategic decisions based on the data. Getting the mix right is discussed in Chapter 8 (in ebook 2).

Almost every organization I have encountered initially underestimated the value of data—in terms of both the value of it to their organization and what it costs to achieve an integrated approach. Wherever you place yourself on Figure 3.8, consider moving two places to the right. That is, if you initially anticipated spending £10,000–£25,000 ($15,000–$40,000) per year on your web analytics, consider a more realistic amount to be £50,000–£100,000 ($75,000–$150,000)—if you wish to achieve what I set out in this book, that is.

Google Analytics Premium may also be the right tool for you (see Chapter 2 for making an informed decision on this). If so, the price tag for the product is $150,000 per year.

Why a GACP Is Important

Google Analytics Certified Partners are companies that specialize in defining and measuring success using Google Analytics (and possibly other complementary tools). Every organization is unique in its metrics strategy, so having an *internal* day-to-day resource within your organization is a key requirement. However, an external GACP resource ensures you apply best-practice principles from the start. That's important, as poor or unreliable data cannot be corrected for later—it has to be discarded.

When it comes to insights, experience counts. A good GACP will work with numerous organizations across a broad spectrum of industries. That lateral thinking is a great asset when it comes to hypothesis building and testing.

SPEAKING OF EXPECTATIONS AND PROCESSES …

When I hear this …	I reply with …
We want to have all this set up by the end of the quarter.	Be realistic with your requirements. If you have a single website domain, with no subdomains, and no e-commerce facility or other complex tracking requirements, then you can build solid data foundations within three months. However, avoid skipping the processes described in this chapter. They are required for good reasons.
As a global company we have multiple country websites, subdomains, and a network of resellers we supply leads to. Our site is aimed at new purchasers, existing customers, investors, and job seekers. We wish to measure and understand all of these areas. What resources should we be considering (budget and staffing)?	This varies in the same way the cost of building your website varies. However, the following is a guide based on my own experience. **Staffing** Three to four full-time staff, one at senior manager level. In addition, build a network of power users to support country offices or specific departments (see Chapter 8, in ebook 2). **Technology** Consider Google Analytics Premium as the potential tool for your organization (see Chapter 2). **Consultancy services** Leverage the experience of experts in Google Analytics. Hire a GACP to support you.[14]
Continuing the above: Who should be our stakeholders for this project?	Your marketing and web development teams are key stakeholders; the project will fail without them. Beyond this, consider any senior manager who has an interest in the digital channel and can contribute budget or resources to the project. Making a contribution to the required investment is a great way to qualify your stakeholders.
We have a junior marketer who can be our internal resource on this project. She can spare one day per week on Google Analytics.	Other than for a small business, this is not enough resources for an organization to make use of their data beyond traffic volume numbers. It will result in data stagnation—repetitive report generation without insight. At the enterprise level, allow a minimum of one full-time person for your

internal resource and use the services of a GACP to support you.[14] See Chapter 8 (in ebook 2).

CHAPTER 3 REFERENCES

1 Analytics maturity models aim to assess an organization's ability to use data in order to move their business forward. A champion of the online model (OAMM) is Stéphane Hamel of Cardinal Path: www.cardinalpath.com/services/online-analytics-maturity-model.

2 Google Analytics refers to its ability to track any Internet-connected device as Universal Analytics. More on this in Chapter 6 (in ebook 2).

3 Professor Hans Rosling is one of my very few modern-day heroes: http://en.wikipedia.org/wiki/Hans_Rosling.

4 The Motion Chart feature for Google Analytics explained: https://developers.google.com/chart/interactive/docs/gallery/motionchart

5 The Google Tag Manager (GTM) is a tool to add and manage page tags such as Google Analytics. It uses a centralized web interface to minimize the IT overhead of tracking: www.google.com/tagmanager.

6 This article from me discusses why and how to implement a fixed exchange rate method when combining multiple currencies into Google Analytics: http://brianclifton.com/multi-currency-support.

7 According to the e-tailing group's 12th Annual Merchant Survey of 2013, 46% of US merchants report a purchase conversion rate between 1.0% and 2.9%: www.e-tailing.com/content/wp-content/uploads/2013/04/pressrelease_merchantsurvey2013.pdf.

8 MarketLive Performance Index, Volume 24, Q1 2014 (PDF download): www.marketlive.com/how-we-think/commerce-strategies.html

9 Baymard Institute compiled 22 different studies containing statistics on e-commerce shopping cart abandonment rates (2006–13): http://baymard.com/lists/cart-abandonment-rate.

10 An excellent article from Steven L. Scott describes the principles behind the multi-armed bandit model: https://support.google.com/analytics/answer/2844870.

11 Specialist client-side testing platforms that integrate with Google Analytics include Optimizely and Visual Website Optimizer.

12 *Don't Make Me Think* is the title of the excellent book by Steve Krug. His simplicity of approach when explaining user experience and website usability was groundbreaking when first published.

13 Econsultancy and Lynchpin. "Online Measurement and Strategy Report 2013": http://econsultancy.com/reports/online-measurement-and-strategy-report

14 Google has an official global network of over 200 certified Google Analytics partners, known as the GACP network: www.google.com/analytics/partners.

4 Assessing Your Data Quality

A major hurdle that prevents analytics data being taken seriously within an organization—by that I am referring to using web data to drive strategic thinking—is the perception that you are looking at noisy, irrelevant, or inaccurate data. A default Google Analytics installation answers only basic questions about your website. That's because without a best-practice setup, all Google Analytics can tell you is the basics: How many visitors came to our site? What pages did they look at? How long did they stay?

Answers to those fundamental questions—and there are many of them—can be incredibly powerful. They allow you to understand how your website performs through the eyes of your visitors. However, from a business point of view, non–data experts in your organization are left with the feeling of "So what?": So what if visitors stay on our website for 3 minutes and this has increased by 10% over the

past 12 months? What does this mean to the business and what should I do with this information?

To the business, such basic metrics are irrelevant. To get past this stage you need to drill down into your data and answer the *business* questions your stakeholders are asking. Drilling down into your data means defining ever-smaller segment sizes. As you do this, the error bars inherent with web analytics get larger (irrespective of the tool used). Without a solid confidence in your data, you can just end up looking at noise or, worse still, similar data points from different reports providing a conflicting story. The analytics project will stagnate.

To minimize the noise and inaccuracy and maintain relevancy, you need to keep on top of your data quality. The process to do this is a health check audit of your data quality. In this chapter I show you how to perform such an audit—without the need to look at code.

Error Bars and How to Minimize Them

As with all data sets, web analytics has its error bars.

- Setup errors—for example, incorrect filter logic applied, or typos in the tracking code. You minimize these with regular health check audits, as discussed in this chapter.
- Incorrect deployment, such as the tracking code missing from certain pages. No tracking code deployed results in no data being collected. You minimize this with regular site scans to determine your tracking code coverage, as discussed in this chapter.
- Visitors using multiple devices, such as desktop, laptop, tablet, and mobile phone. Unless your visitors routinely authenticate—log in to your website—no analytics tool can associate the same visitor across multiple devices. The only way to minimize this is to provide real value for visitors to log in—for example, applying a discount code on your pricing, providing an exclusive members-only area, or providing access to an account representative or a dedicated support area.
- Deletion of cookies. Visitors always have the right to block or delete the Google Analytics tracking cookies. If so, either data will not be collected (visitor blocked your cookie), or your visitor count is over-inflated (visitor deleted your cookie).

  Until the Edward Snowden affair, the deletion of cookies by users was considered relatively low—around 14% per month[1]—and consistent. However, this may now change. You can minimize the problem by having a clear, easy-to-read, and accessible privacy policy on your website. This is not a trivial matter—most are overly long and full of legal jargon there to protect the business rather than the visitor. Getting this right is a key aspect of building trust in your brand. See Chapter 7 (in ebook 2)>.
- Visitors who research online but prefer to purchase at a physical store. This has traditionally been a problem for web analytics tools, as the most important part of the digital trail is lost when the visitor goes offline. However, with Google Analytics' new protocol (Universal Analytics), this can be overcome. See Chapter 6 (in ebook 2).
- Differing processing time frames. A classic example is when payment details received late on a Friday night are processed by your transaction system on a Monday morning. However, Google Analytics will show this as revenue collected at the time of purchase (late on Friday night). There is not a lot that can

be done to minimize this other than to allow for it when trying to reconcile numbers.

DATA QUALITY HEALTH CHECK

To assess your Google Analytics data quality, create a health check audit report. Use this report to examine and summarize aspects of your setup, weight them according to importance, and provide a single numeric representation of the quality of your data—your Google Analytics quality score (*QS*).

The health check audit report provides a succinct summary that captures the information most relevant to the analytics team. It articulates what is working correctly, what is missing from your current data setup, and where the problem areas are; and it lets you focus on which areas to prioritize to meet the organization's data requirements. Figure 4.1 is an example of a first-page summary of the audit report. It's aimed at senior managers in order to get an at-a-glance understanding of the current data quality status.

Scorecard Summary
January 2014: 112,000 visitors, 157,000 visits
Time on site = 3:08; *pages per visit* = 3.75; *bounce rate* = 45%;
AdWords spend = not linked to Google Analytics
Operational since March 2011

		Weight	Status	Weighted Score
1	Account setup and governance	1.0		5
2	Tracking code deployment	1.0		10
3	AdWords data import	1.0		0
4	Site search tracking	1.0		0
5	File download tracking	1.0		0
6	Outbound link tracking	1.0		0
7	Form completion tracking	1.0		0
8	Video tracking	N/A	–	–
9	Error page tracking	0.5		0
10	Transaction tracking	2.0		0
11	Event tracking (non-pageviews)	1.0		0
12	Goal setup	1.0		0
13	Funnel setup	1.0		0
14	Visitor labeling	1.0		0
15	Campaign tracking	1.0		5
	Quality score (*QS*) out of 100			**13.8**

Figure 4.1 *Sample health check audit summary. Ideally, your quality score would be 100, a 100% complete best-practice implementation of Google Analytics. While that should be the long-term aim, obtain a score of at least 50 before you attempt any in-depth analysis of your data.*

In Figure 4.1, first focus on the overall *QS*—the number in the last row. In this example the *QS* = 13.8. This ranges on a scale from 0 to 100, where 100 represents a best-practice setup for your organization. The data quality in the example is not credible for analysis.

The first priority therefore is to achieve a critical value—a score above which

you can perform basic analysis in confidence, which means that the information obtained is based on solid and reliable data. The exact critical score varies by organization. However, to simplify the process I use the following general rule: Until you achieve a *QS* of 50, don't try to do any analysis. A value below 50 leaves too many holes and caveats in the data.

When you can demonstrate a high score—for example, *QS* > 80—you and your peers will have the confidence in the data that you need to base strategic decisions on it.

Be Transparent with Your Quality Score

As your organization will be making key strategic decisions based on your data, you will want to be able to provide your *QS* to back it up. Your *QS* is *the* key metric that shows how reliable your data is. You should monitor it regularly—see the section "Monitoring Your Quality Score" later in this chapter.

Once you understand your quality score, look at the overall scorecard summary. The table summarizes the 15 key items that make up a best-practice implementation by showing the weight (importance), its status (red, yellow, green), and the weighted score for each (weight × status). Clearly you want to have many greens in the Status column and be able to explain the yellow and red items in the report's supporting pages.

Table 4.1 shows the values associated with the status colors.

The weight for each of the 15 items in Figure 4.1 is a relative measure of the importance of tracking that particular item, with 0 representing no importance and 1.0 representing the highest importance. A weight of 2.0 is reserved for transaction tracking, because of its special importance. For each non-green item in Figure 4.1, consider its weight. This is your priority list for work to be done to improve your *QS*.

Table 4.1 *Scorecard Status Values*

Status Color	Explanation	Status Value
Green	Working as expected. No action required.	10
Yellow	Partially working, or not working in an optimal way. Action required.	5
Red	Not working or not implemented. Action required.	1

Page 1 of the audit report (the scorecard summary) is the most impor tant part of the report—as a manager, your focus is on this page. The remaining pages of the report explain why items receive a yellow or red status value.

You can download the full example audit report from http://brianclifton.com/example-audit.

Balanced Scorecard Approach

You may be familiar with the balanced scorecard (BSC) approach—a strategy performance management

technique developed in the 1990s and used by managers to keep track of activities within their control.[2] The health check audit report is based on the BSC technique.

BUILDING YOUR HEALTH CHECK SCORECARD

The health check audit report is built by the analytics team and is typically four to five pages long (including the summary page shown in Figure 4.1). Most, but not all, of the table items will be applicable to you. Mark others as “not applicable” rather than removing them from the table. For example, video tracking (row 8 in Figure 4.1) is not applicable to this fictitious client website. I explicitly state this in order to indicate that video tracking has been investigated and considered not applicable. If that item were missing, it would raise the question, “Have we considered everything?” For large sites with thousands of pages, the precise content and features can easily be forgotten. Therefore, it is good practice to make a formal assessment.

Following the scorecard summary page, the supporting pages expand on why each item received the score it has in bullet point form. For items that receive a green status, no further detail is required. Therefore, the report focuses on the yellow and red tracking items. The information presented should be concise and succinct—that is, in note form. The intention is to help point the person responsible for the specific tracking item in the right direction. Each item should be no more detailed than the following example:

Item 5: File download tracking

- A google.com search reveals we have 5,000 PDF files located in the search engine’s index (they can be found by Google).
- However, no file download tracking is in place.

An alternative could be the following:

- A google.com search reveals we have 5,000 PDF files, 200 ZIP files and 100 XLS documents located in the search engine’s index (they can be found by Google).
- However, only PDF files are currently tracked, and a meaningful filename is not reported.

In both examples, the justification for tracking is stated followed by the current tracking situation.

Weighting Your Scores

As described in Chapter 3, the weighting of items to track is a balance of two considerations:

- Is the activity an important part of the visitor's journey?
- How important is the activity to the business?

The weighting is relative, on a scale from 0 to 1.0, with 1.0 indicating the most important level. There is one exception to this rule—if you have an e-commerce website, weight transaction tracking is 2.0. That is, if you are collecting revenue directly on your website, it is super important to get this tracked correctly in Google Analytics.

For the range of weights, I use five possible values: 0.0, 0.25, 0.5, 0.75, 1.0 (and 2.0 for transactions only):

0.0	Not required. This could be, for example, clicking on images within a carousel. (Why websites bother with carousel images is a mystery to me. The user experience is very poor.)
0.25	Nice-to-know metrics, but the business is not asking for these at present. Examples include clicks on certain links, such as outbound links, or widget usage.
0.5	Somewhat important to know, but not a business priority at present. For example, error messages. These are important if you are launching a new website or rolling out a redesign. But for day-to-day operations, errors are dealt with by the web development team and are usually logged separately to Google Analytics.
0.75	Important metrics though not top priority. For example, these could be social sharing of content while the business is still figuring out its social strategy.
1.0	A key part of the visitor's journey and an important engagement for the business. For example, any action that results in a visitor providing their personal information to you, such as a lead generation form submission.
2.0	Vital metrics. Generally used for transaction tracking only.

The analytics team is responsible for producing the scorecard report. My approach is for the team to make the initial assessment of the weights for each tracking item. They answer the first question: Is the activity an important part of the visitor's journey? This sets the stage for the rest of the organization to understand what your website visitors experience.

The document then becomes the basis for further discussions with your stakeholders—the rest of the business that has invested in the analytics project. The follow-up discussion assesses the second question: How important is the activity to the business? The purpose is to fine-tune and finalize the importance of each weighting. This is a straightforward process and should be accomplished within one or two meetings.

Calculating Your Quality Score

The weighted score is the last table column of Figure 4.1. The calculation is the weight of the item multiplied by its status value (taken from Table 4.1):

$$\textit{weighted score} = \textit{weight} \times \textit{status value}$$

A few examples using Figure 4.1, illustrate how straightforward this is:

Item 1, Account setup and governance: The weighted score is 1.0 × 5 = 5.

Item 2, Tracking code deployment: The weighted score is 1.0 × 10 = 10.

Item 3, AdWords data import: The weighted score is 1.0 × 0 = 0.

Your *QS* is calculated as the sum of your weighted scores divided by the total of the weights representing a perfect score:

$$QS = \frac{\textit{sum of weighted scores}}{\textit{sum of weights} \times \textit{status value for green}}$$

Again, this is straightforward. The *sum of weights* × *status value for green* is the sum of weights multiplied by 10. Using the example in Figure 4.1, this is 14.5 × 10 = 145.0. Therefore, *QS* = 20 / 145 = 13.8.

SCORECARD ITEMS—WHAT TO ASSESS

For each row of Figure 4.1, what needs to be checked, why, and how? Most of this process concerns asking questions about what data is present in your reports, or what settings are configured in your Google Analytics account.

There is no specific order for the items to assess, and this is deliberate. Starting with having a Google Analytics account and deploying the tracking code across your pages (items 1 and 2) makes sense. However, the remaining items are not ranked by importance. This is because when viewing data and tracking code, it is easy for even expert eyes to glaze over. What determines if the tracking works correctly or not can be subtle. The random ordering of items in the scorecard forces the assessor to crosscheck reference points and is therefore a self-check for such subtleties.

For this reason, I recommend that only one person be responsible for the status assessment. They should be comfortable with HTML, JavaScript, and the architecture of your websites. This person should be on the analytics team, not someone from IT.

Aren’t These Tracked by Default?

Regardless of how a visitor arrives at your website, as long as you have deployed your tracking code, Google Analytics will track by default the visit, the visitor (the same visitor may make several visits), and the pages viewed by your visitors. That is a simple yet powerful data set that allows more than 100 visually rich reports to be automatically built for you. These show

- The time and date visitors arrive, and from which countries and cities—accurate to approximately a 25-mile (40 km) radius.
- The number of times they have come: are they a first-time visitor or a repeat visitor, and if so, how many times did they come before?
- What drove them to visit you: which search engine did they use, which social website or other website was involved in bringing you your visitors?
- The content visitors view: which pages are popular, how much time is spent on each page and on your entire website?

However, counting how many visits, where the vistors came from, and what pages they looked at is still a small piece of the data pie. It tells you almost nothing about engagement—whether your visitors achieve anything worthwhile to them, or how close they get to becoming a customer or repeat customer. It tells you nothing about your visitor value—that is, are you receiving high-value visitors or low-value ones? And it tells you nothing about which pages are important to your visitors—the pages that are most influential or persuasive—versus pages that just waste your visitors’ time.

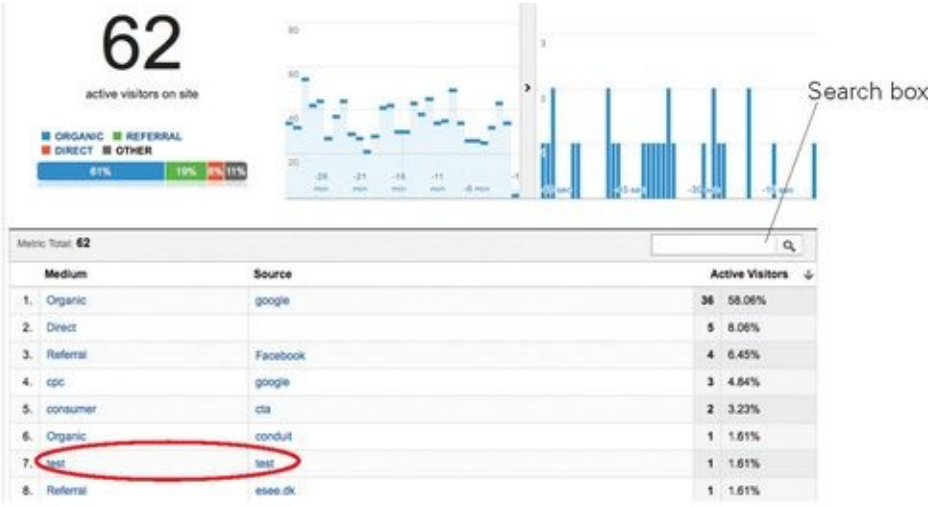
Items 3 through 15 of the scorecard cover the collection of data points required to get a more informative picture of your visitors’ experiences.

1. Account Setup

The first row of the scorecard summary table, Google Analytics Account Setup, concerns your Google Analytics account governance and includes the following three areas:

Check in Real Time

The real-time reports of Google Analytics are a subset of reports that show you what's happening on your site—*as it happens*. Visit data is processed within a few seconds of the visitor's arriving on your site. Here is an example.



A real-time report showing an up-to-the-second view of where visitors are coming from

I use real-time reports to test what data is being reported on. The technique is to show real-time data for only your *own* visit—that, is while you are auditing your website. In this way you can see what data is being sent to Google Analytics (if any) that corresponds to your action.

Open two browser windows—one for your Google Analytics real-time reporting, one for your website. To isolate yourself from all other visitor traffic, visit your website with the addition of two extra campaign tracking parameters in the URL. For example, www.example.com?utm_source=test&utm_medium=test will be displayed in your real-time reports, as shown in table row 7 above. If you do not see the test row in your table, use the search box. By clicking on *test* in the Medium column (row 7), you can automatically segment your data in Google Analytics. That is, you isolate just that visitor traffic. As this URL is unique to you, it corresponds to your actions.

Now you can navigate the real-time reports and only view your own data. In your other browser window, click around your website viewing different pages. Assuming you have the basic Google Analytics tracking code in place, you will see the page URLs you have clicked on appear in the real-time content report.

Determining whether an action is tracked at all and how is the basis for many of my tracking tests within the scorecard.

Account Structure

Google Analytics is not a relational database. You cannot simply throw data into a big pot and then figure out the data relationships later. How you collect your data determines its structure. You can manipulate your data structure within your

account setup using filters.

What to assess:

- Are you excluding visits from your own staff? These visits could represent a significant proportion of your total traffic, and they are not from your target audience. If tracking your own staff visits is important, create a separate report set for them.
- Are you excluding visits from your third-party agencies?
- Is there a backup report set to protect you if something should go wrong with your data?
- Is there a testing report set allowing you to experiment with changes before applying them to the main data set?
- Are any content filters applied and, if so, are they valid and working?
- Is the number of report sets (referred to as *views* in Google Analytics) reasonable and manageable for your organization?
- What other filters are present that are manipulating the data?

Where to check: In order to make your assessment, you need admin access to your Google Analytics account. Within the Administration area (Figure 4.2a), check what views are available and what filters (Figure 4.2b) are being applied to your data.

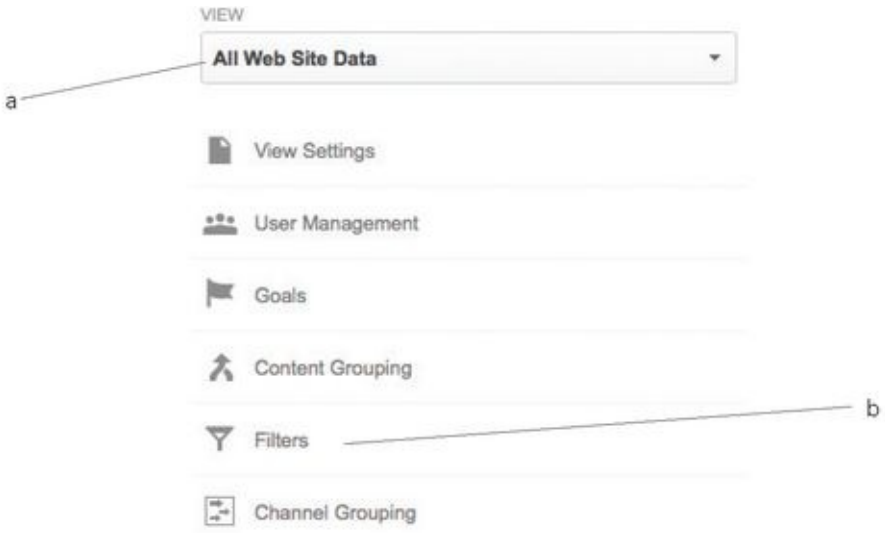


Figure 4.2 *The Google Analytics Administration area for a report set: (a) assessing different report sets (views); (b) assessing applied filters*

People and Access

Who has access to your data? This includes people both in your organization and in external third parties (agencies, consultancies, developers, partners, resellers, affiliates, and so forth). Similarly, who has the right to change the setup? As is the case with your bank account, your data is confidential and should only be viewable by authorized people.

What to assess:

- How many people have access to your data?
- What level of access do people have? View-only access, or full admin rights to change anything?
- How many admin users are there? This should be limited to a very small subset of users who have experience with Google Analytics administration and take full responsibility for their actions. The total number of account administrators should be as small as possible, and the requirement of being an administrator should be justified for each individual. As a rule of thumb, I red-flag any account with more than ten administrators.

Where to check: To make your assessment, you need admin access to your Google Analytics account. Within the Administration area shown in Figure 4.2, select the User Management menu item and review the people listed and their associated permission levels.

Content and Information

Documenting your setup (building a health check audit report is part of this), and understanding what type of data is being collected, is an important part of good governance. For example, Google Analytics has some restrictions on the type of data and volume of data you can send it. These are covered in its terms of service (US version here: www.google.com/analytics/terms/us.html[3]).

What to assess:

- What is the volume of data being sent to Google Analytics? This includes all pageview, event, and transaction data. The free version of Google Analytics is limited to 10 million data hits per month. For Premium users, the standard limit is 1 billion data hits per month.
- Is personally identifiable information (PII) being collected? This is a red flag. No PII can be reported in Google Analytics. If it is, you risk your account being closed down. (*Note:* I am specifically referring to the collection of PII by Google Analytics. Collecting visitor PII in your contact database is fine, so long as this is done with the consent of your visitors.)
- Is there any documentation explaining the data collection methodologies employed? That is, has a scorecard audit been delivered previously, and does an implementation guidelines document exist (see Chapter 3)?
- Is there a suitable privacy statement on your website explaining how your organization collects and processes visitor data?

Where to check: To assess your total data volume, access your reports and enter the Behavior section (Figure 4.3). The overview report (a) lists your pageview volume for the past 30 days—the default date window. Add this number to the total number of events tracked, obtained from the Events ⇨ Overview report (b).

This number should be less than 10 million for a 30-day period in order to stay within the free Google Analytics terms of service. If not, consider the options discussed in Chapter 2 for this scenario.

Figure 4.3 *The Google Behavior report area: (a) pageviews; (b) events; (c) page content*

PII can end up in your Google Analytics reports a number of different ways—for example, via transaction tracking, custom dimensions (known in older versions of Google Analytics as custom variables), event tracking, and pageview tracking. Any method of collecting PII breaks the Google Analytics terms of service. If someone has consciously made the decision to track PII in your Google Analytics account, then your situation is pretty dire. You must close the offending web property and all report sets for it, delete it (losing all previous data), and start again without collecting PII.

Even without making a conscious decision to track PII, it is possible to inadvertently collect it using Google Analytics—see the sidebar "How PII Gets into URLs." To assess this, review your Site Content ⇨ All Pages report (Figure 4.3c). This report contains a list of all page URLs viewed by your visitors. Conduct a table search for any possible PII collected in the visited URLs. Typically this is for potential field names such as "name," "email," or "address." Searching for the email @ symbol is also a good way to bubble up any potential email addresses captured.

If PII is present in your reports, you must close the offending web property, delete it, and start again without collecting PII.

How PII Gets into URLs

If you collect personal information via a web form, ensure your developers use the POST method for form submission. This ensures the information contained in the form is not visible in your page URLs (for best-practice privacy reasons, you should also send the information encrypted via https).

If form submission data is not transmitted via the POST method, any form field information is appended to the page URL as a query parameter. The page URL is what Google Analytics tracks by default. Hence if PII is present in your URLs, this will appear in your Google Analytics reports. An example URL could be /form/subscribed.php?name=Brian%20Clifton&email=brian@brianclifton.com

Apart from bad practice, losing the trust of your visitors, and breaking the Google Analytics terms of service, passing around personal information on the Internet in clear text form is likely to get you into trouble with the data protection authorities of the countries you operate in.

See Chapter 7 (in ebook 2) for a detailed discussion about PII and privacy.

2. Tracking Code Deployment

The Google Analytics tracking code (GATC) is the JavaScript snippet of code (typically a dozen or so lines) that you paste into the header area of all your pages. Without this tracking code present, no data collection can take place. In addition, a patchy deployment of the GATC—that is, some pages tracked, some not—leads to unexpected results, often the double counting of visitors.

What to assess:

- Is the tracking code working—that is, collecting data?
- Is cross-domain tracking required and implemented? This is when a visitor traverses more than one website that you own. This can be a subdomain (www.example.com ⇨ blog.example.com) or a full domain change (www.product-site.com ⇨ www.support-site.com).
- What version of the tracking code is running? Since the launch of Google Analytics, there have been three versions of the tracking code. The latest, launched in late 2013, is named analytics.js. It is commonly referred to as Universal Analytics.
- How is the tracking code placed on your pages—using code snippets or via a TMS? If a TMS, which vendor and who is responsible for it?
- What proportion of your website pages can be tracked? Your target is 100%. That is, at the very least your GATC should be deployed on *all* of your public-facing website pages. After all, if your organization has gone to the trouble to produce content, you should be tracking it. Otherwise, consider culling the page.

Where to check: Checking the GATC is the most technical part of the health check audit. To perform the check, you could manually review your web page HTML source code on a sample of pages. However, that is a laborious task and is prone to human error. Instead, there are a number of tools that can simplify the task and that offer ways to automatically detect errors and issues for you.

Google Analytics Debugger—official Google Chrome extension Loads the debug version of the GATC for all websites you browse. It prints useful information to the developer console that can tell you when your analytics tracking code is set up incorrectly. It provides a detailed breakdown of each tracking beacon sent to Google Analytics. Free to use. (I use this daily!)

https://chrome.google.com/webstore/search/google-analytics-debugger

Google Tag Assistant—official Google Chrome extension Helps to troubleshoot the installation of various Google tracking codes on your website—verifying you have installed beacons correctly as you navigate pages. It reports which beacons are present, reports errors found, and suggests implementation improvements. Free to use.

https://chrome.google.com/webstore/search-extensions/tag-assistant-by-google

Analytics Debugger—third-party website service Can scan and debug your website implementation (and uniquely your mobile apps) for numerous tracking beacons, including Google Analytics. In addition, it can monitor for tracking code changes and provides *interactive* debugging—showing you what data is sent as you interact with your pages. Freemium tool. www.analytics-debugger.com/brianclifton

Web Analytics Solution Profiler (WASP)—third-party Chrome extension Enables you to test and debug your website tracking beacons from virtually any tracking tool on the market. It provides a visualization of scripts and tag dependencies based on an auto-detecting algorithm. Freemium tool. www.webanalyticssolutionprofiler.com

Tag Inspector—third-party website service Intelligent crawler to scan your website. Hierarchy views give you a complete picture of your tracking beacons—whether in the source code of your page, within a tag management system, or piggybacking off another tag. Freemium tool.

http://taginspector.com/brianclifton

3. AdWords Data

AdWords, Google's advertising platform, has its own separate reporting interface. However, this only allows you to see visitor click-throughs to your website—the number of clicks on your ads, not all the things visitors do once they arrive on your website. Clearly, there is an advantage to following the visitor's journey.

Doing so requires linking your Google Analytics account and your AdWords account. Once this is done, not only will you be able to view the full visitor journey (from ad click-through to engagement to conversion) but you can also import your AdWords cost and impression data into your Google Analytics reports.

What to assess:

- Can Google Analytics identify your AdWords visitors?
- Is your AdWords cost and impression data being imported?
- If you use more than one AdWords account, do your payment currencies— the currencies you pay for your ads—match? To avoid confusing the performance of your campaigns, you will want to avoid mixing AdWords currencies within one set of reports.

Where to check: All visitors who click through from your AdWords ad will have an extra parameter (*gclid*) appended to your landing page URLs, for example, www.example.com/product/?gclid=Q76WERbcTYuAWw. Therefore, the first check is to do a search on Google that displays your ad, click on it, and inspect the URL of your landing page. If the gclid parameter is not present, Google Analytics will not identify it as an AdWords visit. Instead, it will regard the visit as coming from Google organic search (free search)—a very different system! If you notice this problem, contact the person responsible for your AdWords advertising and ensure your AdWords account is linked to your Google Analytics account.[4]

Assuming the gclid parameter is present on your landing pages, check the AdWords section of your Google Analytics reports for cost and impression data (Figure 4.4). If these are present, then both the reporting of AdWords visitors and the concomitant cost and impression data imports are working. If not, contact the person responsible for your AdWords advertising and ask about the linking of your AdWords and Google Analytics accounts.[4]

Figure 4.4 *Checking AdWords data*

4. Site Search Tracking

Site search is the terminology used to describe your website's internal search engine. For any website with even a moderate number of pages, site search is a critical part of your visitors' website experience. A good site search facility can make the difference between a visitor having an efficient, enjoyable experience and a painful one that can cost you business.

The queries (search terms) your visitors enter are a great visitor feedback mechanism. Your visitors are telling you exactly what they are looking for, in their own language and using their own terminology. Analyzing query terms can prove insightful when it comes to understanding how your visitors think about your products.

What to assess:

- Can Google Analytics detect usage of your internal site search facility?
- If this usage is tracked, are the keywords a visitor uses reported?
- Are keywords categorized into sensible groupings? For example, product search, support search, store locator search.
- Are you tracking zero search results—instances when your site search facility returns no results back to the user?

Where to check: Site search usage has its own dedicated set of reports within Google Analytics (Figure 4.5). Review these reports. Tracking zero results is particularly insightful. Typically a separate site search category is used for this. Therefore, check for its presence.

Figure 4.5 *Checking your site search data*

5. File Download Tracking

By definition, a file download is not an HTML file, and therefore this cannot receive a GATC within it. Therefore, the technique for tracking these is to trigger a Google Analytics tracking call when a visitor clicks on a download link.

What to assess:

- Are file downloads being reported?
- Are reports clear on what file has been downloaded? Do they make sense to your marketing team, or is a lookup table required?
- Can you differentiate file types—PDF, DOC, ZIP, and so on?

Where to check: There are two ways to track a file download—either as a *virtual pageview* or as an *event* (see the appendix). As a quick assessment, check your real-time reports while browsing your website. Click on a file download link and view both the content and events reports within the real-time section. This will reveal if any tracking of file downloads is in place and, if so, what method is being used.

Regardless of the method used, file downloads should be categorized in a logical manner—for example, a category name of *pdf*, *downloads*, or *files* (event tracking), or */downloads/pdf* (virtual pageview tracking). The exact naming is not important—as long as it is logical to the report user. For good account governance, ensure you document your specific setup before assigning a status value of green for this item.

6. Outbound Link Tracking

Outbound links are links on your website that send a visitor away to another website. This can be a completely different, though relevant, destination— such as to a reseller or partner website. The link could also send a visitor to another web property that you own—for example, a link leaving your product website could take the visitor to your support website.

As a click on an outbound link is taking the visitor away from your website, that action is not tracked in Google Analytics by default. Therefore, the technique is to trigger a Google Analytics tracking call when a visitor clicks on an outbound link.

What to assess:

- Are click-throughs on outbound links tracked?
- If so, how are they tracked?
- Can you differentiate your different types of outbound links—to websites you own, to resellers, and so forth?

Where to check: As with tracking file downloads, outbound links can be tracked as a virtual pageview or an event. Therefore, detecting whether these are reported is the same process. Check your real-time reports while browsing your website. Click on an outbound link and view both the content and events reports within the real-time section. This will reveal if any tracking of outbound links is in place and, if so, what method is being used.

Regardless of the method used, outbound links should be categorized in a logical manner—for example, a category name of *outbound*, *links*, or *resellers* (event tracking), or */external/site-name* (virtual pageview tracking). The exact naming is not important—as long as it is logical to the report user. For good account governance, ensure you document the setup used for outbound link tracking before assigning a status value of green for this item.

7. Form Completion Tracking

When a visitor submits a contact request form, or any other type of form (subscription sign-up, registration), is the submission tracked correctly?

What to assess:

- Is it possible to differentiate between a form view and a form submission?
- Does Google Analytics collect any PII?

Where to check: In most cases, you will want your forms and form submissions tracked as pageviews. That is, the visitor sees the initial form page, and if they submit their details they then view a thank you or confirmation page. Try this yourself on your own site. If the URL does not change when you submit your form, you will not be able to ascertain how many submissions were made or calculate your form conversion rates.

The best way to verify this is within your real-time reports. While browsing your website, view your form, submit it, then check both the content and events reports within the real-time section to see if any pageview or event data is sent. If you have numerous forms on your website, ensure a logical naming structure is employed—either as a virtual pageview path or as an event category.

While checking if form views and submissions are tracked in Google Analytics, review if any PII is captured. This was discussed earlier in this section for item 1 of the scorecard, Account Setup. Doing so again here is a good secondary check. After all, collecting PII breaks the Google Analytics terms of service and means you will have to close and delete *all* previously collected data for a web property.

8. Video Tracking

Moving beyond film and animations, video is gaining popularity as a method to demonstrate products, provide how-to guides, troubleshoot problems, and record time-sensitive events, such as conference presentations. If you have embedded video on your website, you will want to know if people are watching the content and to what extent.

What to assess:

- Is Google Analytics tracking whether visitors play an embedded video?
- Is Google Analytics tracking whether visitors play an embedded video and view *x*% of it? For example, 50% is a good sign of engagement.
- Is Google Analytics tracking whether visitors played an embedded video to completion? Clearly, watching to the end is an indication of a very strong engagement.

Where to check: In most cases, you will want your video interaction tracked as an event—an interaction within a page. Therefore, check for this in your real-time reports while browsing your website. Click on an embedded video file to start the play. At the same, time view your events reports within the real-time section. This will reveal if any tracking of video interaction is in place. At a minimum, you should see an event reported for the start of the video play. Other desirable metrics include *played to completion* and *played x%* (for example, *played 50%*).

If tracking is in place, ensure a logical naming structure is employed, so you can easily identify what video was watched and view it separately from other types of events. As there are numerous ways to define a visitor's engagement with a video, document your setup before assigning a status value of green for this item.

9. Error Page Tracking

Generally, error pages are tracked by default. If you have deployed the tracking code to your page templates, an error page such as “page not found” will also be tracked. The problem is that the default tracking behavior results in the error being tracked as a regular pageview, with no indication it is actually an error.

What to assess:

- If the visitor has encountered an error with your site, is the error information sent to Google Analytics?

Where to check: Check this in your real-time reports by viewing an obvious error page. For example, enter the following URL into your browser: www.example.com/product-mytest (replacing example.com).

When viewing the content report shown in your real-time section, you will most likely see */product-mytest* showing as a valid pageview. If so, you will notice there is nothing to indicate that this page is in fact generated by an error (a 404 page). This should be corrected by using a virtual pageview call to Google Analytics instead. If you have this in place, you will see something similar to */error 404/product-mytest* in your Real-Time ⇨ Content reports. The exact naming convention is not important as long as it is clear to the report user that the pageview is generated by an error.

Alternatively, check your page titles within the content section of your reports. If page titles contain information that a page URL is in fact an error, you can use this instead to group and highlight all URLs that are generating errors to your visitors. Document which setup (if any) you have in place for identifying errors before assigning a status value of green for this item.

10. Transaction Tracking

Transactions by their nature have many moving parts. From a technical viewpoint, tracking a transaction can be quite difficult. Cross-domain tracking is often required for a visitor to complete their purchase— tracking a visitor from www.example.com to shop.example.com, or www.example.com to www.payment-gateway.com and back to www.example.com.

In addition, there are numerous ways a product can be purchased: the initial purchase; modifying the purchase at a later date (paying more); changing the delivery details at a later date (not paying more). Which of these should be tracked as a transaction requires careful consideration. Generally I recommend only the initial purchase be tracked, as this is the only action that directly relates to the performance of your website—its marketing, its content, and its usability.

What to assess:

- Are transactions tracked on your site correctly? That is, do the numbers correlate with your back-end system that processes the orders?
- Is cross-domain tracking implemented and correct?
- Are multiple currencies taken into account? For example, combining USD with EUR with GBP is not desirable.

Where to check: Real-time tracking of transactions is not currently available in Google Analytics. Therefore, you will need to check the standard reports for these —Conversions ⇨ E-commerce reports (Figure 4.6).

As an initial sanity check, select five or so products from your product performance report and review the quantity purchased, unique purchases, and product revenue numbers. Do these match your company system that processes orders? The match is unlikely to be exact for a variety of reasons: order errors, returns, time differences when an order is received versus when it is processed, and so forth. As a rule of thumb I accept an error of up to 5%. If the errors are larger than this, then there is likely a problem. Perhaps different currencies are being mixed or there is a technical error with the implementation?

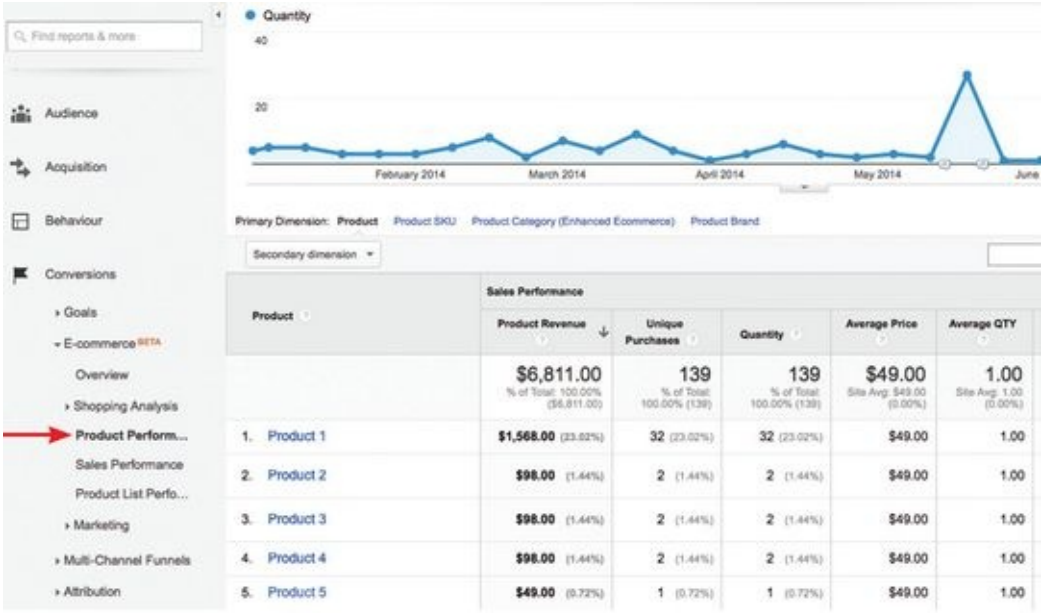


Figure 4.6 *Example of a transaction report*

To shed more light on this, conduct several different test transactions yourself—transactions where you know exactly what you bought, the quantities, and their costs. On completion of your purchase, the transaction ID of your order, displayed to the purchaser, can also be found in the E-commerce ⇨ Transaction report. Within this report, click on your transaction ID and view your specific order for correctness.

11. Event Tracking

Events are in-page actions that visitors complete that do not result in a pageview. That is, they are not tracked by default in Google Analytics. Some of these you will have already covered in items 5, 6, and 8. Other potential events include clicks on internal advertisements; rotating through a carousel of images; moving around a long page of content using HTML bookmarks; clicking on a click-to-call button; adding, removing, or modifying a shopping cart item; and interacting with widgets, such as a loan calculator.

What to assess:

- A key assessment to be made is whether any event tracking is required at all. Not every non-pageview click a visitor makes has to be tracked—that would generate a large amount of noise. Carefully assess what has to be tracked. For example, I have found little purpose in tracking carousel images, other than to show that no one is interested in them!

Where to check: Check your event tracking reports (Figure 4.3b) to see what data is being collected and decide if it is valuable to your organization. If these events are required, ensure that what they define and what triggers them is documented before assigning a status value of green for this item.

12. Goal Setup

A *goal* is a key engagement point that indicates some kind of success—the success of converting an anonymous visitor into something more. An obvious goal is the conversion to a customer (purchase). However, in most scenarios it is the strengthening of the visitor relationship—either anonymously or for an existing customer. For example, these can be the download of a PDF catalog; submission of a contact request form; subscription to a newsletter; using your store finder; sharing your content socially; or adding a review or comment to your site.

What to assess:

- Check your conversion reports to establish whether any goals have been set.
- If so, do these reflect the goals of the business—that is, are they relevant?
- Have your goals been monetized?

Where to check: If your Conversions ⇨ Goals ⇨ Overview report contains any data, then you have at least some goals configured (Figure 4.7). Otherwise, you will see a message with words to the effect “This report requires goals to be enabled for the view.” If data is present in this report, review the relevancy to your business. Are the goals tangible indicators of success? Before assigning a status value of green for this item, ensure an implementation document is available explaining what goals are defined and what triggers them.

Figure 4.7 *Sample goal conversion overview report*

If you know what goals *should* be defined in your reports, you can verify that these are working by monitoring your real-time reports. Visit your website and complete a goal conversion yourself; then view the Conversions report shown in your real-time section.

☛ *Goals are synonymous with KPIs. For more details, see Chapter 9 (in ebook 2).*

13. Funnel Setup

A funnel is a well-defined sequence of steps that a visitor goes through to achieve a goal (as in item 12). In this context, *well-defined* means that the visitor must go through the sequence of steps in order to reach the goal. A classic example is a checkout process: the visitor must go through certain steps in order for a successful transaction to take place. The steps may consist of add to cart ⇨ enter delivery details ⇨ pay ⇨ receipt page. The receipt page is the goal in this scenario (the confirmed purchase), and the preceding steps are the sales funnel. You define these steps in Google Analytics so that you can visualize shopping cart abandonment rates, pain points, and so forth.

Non-transactional sites also have funnels. A contact request (or subscription) form is always a two-step funnel—the initial viewing of the form followed by its successful submission. However, not all goals have an associated funnel. Typically a visitor can easily download your PDF literature without a funnel process. You could of course force a visitor to register for your downloads, thereby creating a funnel. However, avoid this. People will not tolerate having to register for your brochure or a support document—something all users expect without hindrance (in the same way people will not accept hindrance walking into your Main Street store). Registration should only be used for high-value content, as perceived by your visitors, not by your business.

What to assess:

- Check your conversion reports to see if any goal funnels have been set.
- Check whether the data drop-off through the funnel appears sensible.

Where to check: If your Conversions ⇨ Goals ⇨ Funnel Visualization report contains any data, then you have at least some funnels configured (Figure 4.8). Otherwise, you will see a message with words to the effect “This report requires goals to be enabled for the view.” If data is present in this report, review the relevancy to your business.

Figure 4.8 *Example of a funnel visualization report*

14. Visitor Labeling

Visitor labels are labels Google Analytics can attach to your visitors' activity. The Google Analytics terminology is *custom dimensions* and *custom metrics*. However, I prefer the nontechnical term *label* to describe these. (Before the Universal Analytics update of 2013, these were referred to as *custom variables*).

There are multiple ways to use labeling. The most obvious use is to distinguish between a "customer" and a "noncustomer." A visitor may have purchased from you only once but may have visited your site many times. The same technique can be used for a "subscriber," a "contactor," or any other visitor who can be identified as engaged or interested. It is a powerful method to use for segmentation.

What to assess:

- Check your reports for visitor labels in use (*custom dimensions* or *custom variables*).
- If in use, do they reflect high-value actions on your site—such as becoming a customer?

Where to check: If you are using the latest Universal Analytics tracking code, check whether any custom dimensions have been configured; review your web property Administration section, as shown in Figure 4.9. In this example, there is one *custom dimension* defined, "AVP remembered."

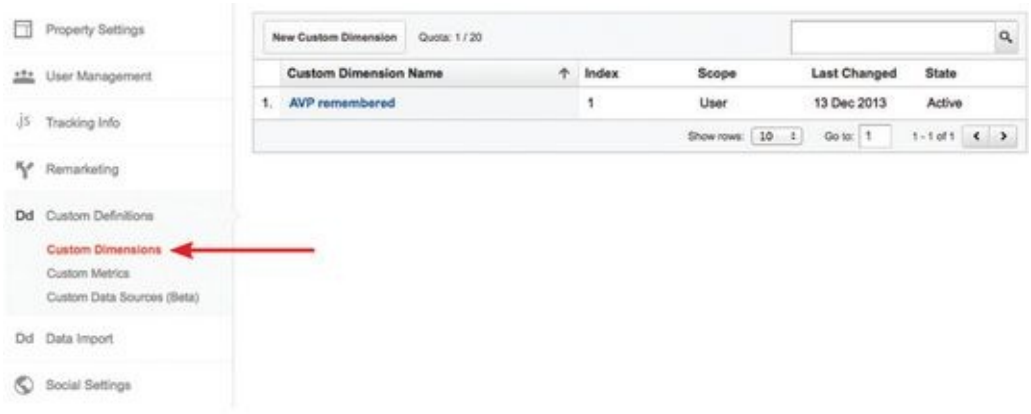


Figure 4.9 *Viewing visitor labels for a Universal Analytics deployment*

If you are not using Universal Analytics (see scorecard item 2), custom variables can be viewed in your Audience ⇨ Custom ⇨ Custom Variables report. There are five slots available for custom variables. Check all five keys for data.

Regardless of which tracking code deployment you have in place, a good implementation document is required in order to understand what your labels are defining and what triggers them. Ensure this is in place before assigning a status value of green for this item.

15. Campaign Tracking

Are your campaigns being tracked? If you are running paid search advertising (such as AdWords), banner ads, email marketing, social media marketing, affiliate marketing, even offline marketing, you will want to know which specific ads and which specific campaigns are working for you. To track these you must modify your landing page URLs by adding parameters to them. This is called *campaign tracking* and is described in Chapter 6 (in ebook 2).

Setting up campaign tracking is straightforward. However, for reports to make sense they must be implemented and done in a methodical way. *Note:* Tracking AdWords visitors is covered in scorecard item 3.

What to assess:

- Aside from AdWords, are other campaigns tracked?
- If so, are campaigns logically named and relevant to your marketing communications department?
- Are campaigns correctly grouped by channel?

Where to check: Your acquisition section contains two key reports for checking your campaign tracking. From the All Traffic report, select the primary dimension of *medium* and review the data table as shown in Figure 4.10. The four rows shown, representing the mediums *none*, *organic*, *cpc* (AdWords), and *referral*, are the default channels that Google Analytics detects. If you only have these four mediums present in this report, then campaign tracking is not set up. Optimally, the number of different mediums tracked should be around ten. More than that can indicate an issue with your setup, such as the misalignment of campaign tracking naming terms.

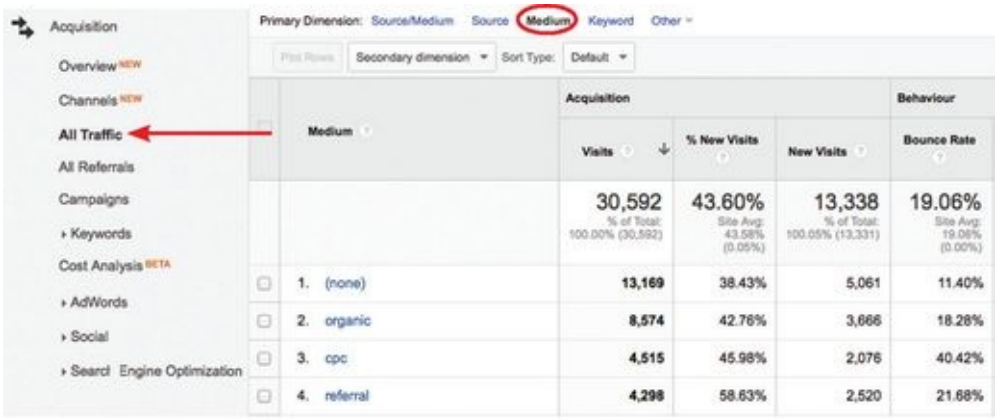


Figure 4.10 *Sample mediums report showing which channels are sending traffic*

Review your Acquisition ⇨ Campaigns report. This lists all your campaign names, including any from AdWords. Are these logically named, and is your marketing team able to understand them at a glance?

For campaign tracking, Google Analytics reports WYSIWYG—what you see is what you get. Anyone (even outside your organization) can create URLs to your website with campaign tracking parameters appended to them. If they are

gibberish, contain typos, or are simply wrong, they will show up in your reports as written. Document how your campaign tracking is defined so that reports can be understood and filtered if necessary. Ensure this documentation is in place before assigning a status value of green for this item.

Default Referrers in Google Analytics

As long as the tracking code is deployed across all of your pages, Google Analytics tracks *all* visitors to your website by default, regardless of where they came from. However, unless you tell it otherwise, Google Analytics can only assign a visitor to one of three* channels and none of these contains campaign-specific details:

- Organic—a visitor who has conducted a search engine query and clicked through to your website via a non-paid listing (that is, not an ad)
- Referral—a visitor who has found a link to your website on another website and clicked through (an affiliate site, partner site, trade association listing, or the like)
- Direct—a visitor who has typed your web address directly into their browser because they remembered it (or used their existing browser bookmark)

Clearly, only reporting three channels is limiting. The fix is to deploy campaign tracking, which enables you to track unlimited campaigns in any way you wish.

* There is a special fourth case for social network visits that Google Analytics will automatically assign—visits from Facebook, Twitter, Google+, and similar sites. However, campaign information for these social sites remains unavailable unless campaign tracking is deployed.

An important consideration when tracking social visitors is that at least half will come via a mobile app,[6,7] not via a standard web browser. Such referrals cannot be detected by Google Analytics or any other tracking tool. This means the name of the social site (Facebook, Twitter, LinkedIn, Google+, and so forth) cannot be detected. However, this can also be overcome by the use of campaign tracking.

MONITORING YOUR QUALITY SCORE

Your Google Analytics quality score is an important metric in itself—your data health check that is part of any analysis. For example, if there is a sudden unexpected change in visitors clicking on a marketing campaign, the analyst team will first refer to the overall quality score and the scorecard detail that comes with it to assess the validity of the sudden change: is the campaign in question tracked correctly?

Determining your quality score is not a one-time, set-and-forget process. Data quality can and does degrade over time. There are many reasons, but it comes down to the rapidly evolving medium of the web. Commercial websites are in a constant flux. New content is created; existing content is updated; marketing techniques evolve; new technology and new ways of presenting information come to the fore. Your business is also evolving, with new and updated products, staff changes, events, PR, policy changes, new services. After a year, if you compare your website before and after, a great deal will have changed, even if the look and feel remain the same.

This constant flux of change affects your data quality—for the worse. Tracking codes can go missing, errors get introduced, new tracking methods don't follow the original blueprint, and so forth. For large organizations, different departments working closely together at the start of the data project can drift away from the central approach and focus on their own silos. This can affect the data for all your Google Analytics users. For all of these reasons, it is important to regularly audit your Google Analytics implementation with a scorecard review.

How often should you audit?

This will depend on how much flux your website is in. To keep things manageable, I recommend at most once per month and at least once per year. Going through the audit process may sound painful, but after your initial scorecard audit, updates usually happen as part of the regular analysis of reports. That is, as part of their ongoing analysis, the analysis team will be constantly validating data. Anything found missing should be reported and the quality score adjusted for that period. The fix can then be prioritized for your next website update.

Figure 4.11 is a dashboard approach I use for monitoring a website's Google Analytics quality score over time. It shows how much confidence the analytics team has in the current data set and in which direction the quality score is heading.

Figure 4.11 *Google Analytics quality score tracked over time*

SPEAKING OF DATA QUALITY …

When I hear this …	I reply with …
Our stats reports in Facebook.com show much greater numbers than what Google Analytics reports. Which is right?	Unless campaign tracking has been used on your landing page links (the links on Facebook that point back to your website), visitors from Facebook will be significantly undercounted by Google Analytics. This is because mobile app users will not be correctly tracked. In addition, campaign tracking parameters are required to report campaign-specific information.
Who should be assigned to building and maintaining our Google Analytics scorecard?	The responsibility for this should lie with the analytics team working in conjunction with the web development team. Typically the person conducting the scorecard audit is an analyst—a person expert in Google Analytics and comfortable viewing HTML source code and JavaScript.
Who should be assigned to building and maintaining our campaign tracking?	This should be owned and maintained by your *internal* marketing team. Even though you may use external agencies to place ads and build marketing campaigns for you, the tracking should be centrally coordinated within your organization. This ensures a consistency of tracking across the broad spectrum of marketing activities—both digital and non-digital.
Our quality score varies each month—sometimes as high as 90, sometimes as low as 70. Should we be concerned?	Your quality score will vary over time due to the natural flux of all things digital. However, ensure your team understands why there is variance and try to minimize it. Learn from past changes. Also be transparent with your score items so that other teams don’t waste time on analysis that may be flawed or lose trust in the reporting method. Even a negative change may be viewed in a positive light if you are transparent and can explain it.

CHAPTER 4 REFERENCES

1 Empirical studies using real data for cookie deletion rates are notoriously difficult to conduct. Most reports are survey-based studies, which I have concluded are inaccurate (in my experience, survey respondents exaggerate their vigilance when it comes to sensitive subjects such as privacy). However, a detailed empirical study was conducted by Paul Strupp and Garrett Clark in 2009 (both formerly of Sun Microsystems). The original article is no longer available, but the Internet Archive does contain a copy: https://web.archive.org/web/20090828151603/http://blogs.sun.com/pstrupp/entı

2 The balanced scorecard approach is defined at http://en.wikipedia.org/wiki/Balanced_scorecard.

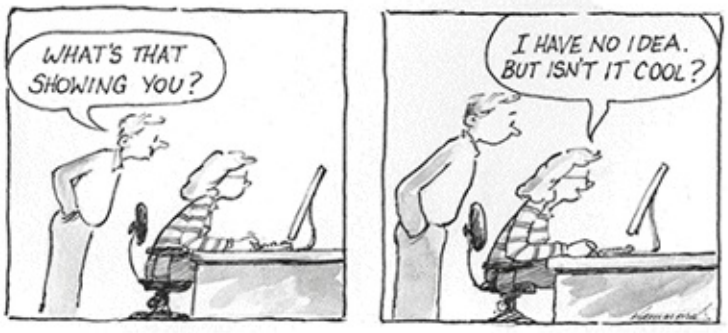
3 The Google Analytics terms of service vary by country. Ensure that you view the correct version relevant to you by selecting the Terms of Service link at the bottom of your Google Analytics reports.

4 Step-by-step instructions for linking AdWords with Google Analytics: https://support.google.com/analytics/answer/1033961

5 In July 2014, according to a compilation of statistics by Statistic Brain, 60% of tweets came from third-party applications: www.statisticbrain.com/twitter-statistics.

6 From Facebook's quarterly earnings slides (Q2 2013), 71% of Facebook users access the site via a mobile device: http://www.cnet.com/news/facebook-earnings-by-the-numbers-819m-mobile-users/.

5 Jumpstart Guide to Key Features



This chapter is a quick overview of report features so you can see how to use Google Analytics to drive your business forward. This is not an anatomy of the user interface or an endless roll of report screenshots. By the end of this chapter, you should have the confidence and interest level to investigate further. I use Chapter 10 (in ebook 2) to illustrate how some of these features are used to provide real-world insights.

REPORTS OVERVIEW—WHAT'S AVAILABLE

While this book focuses on the significance of the available reports from a website owner's point of view, the new Universal Analytics protocol allows you to collect data from anything connected to the Internet ("the Internet of things"[1]), so these reports can be applied in other ways. The significance of Universal Analytics is discussed in Chapter 6 (in ebook 2).

Figure 5.1 shows the Audience Overview report. This is the initial report that is loaded when you log in to Google Analytics.

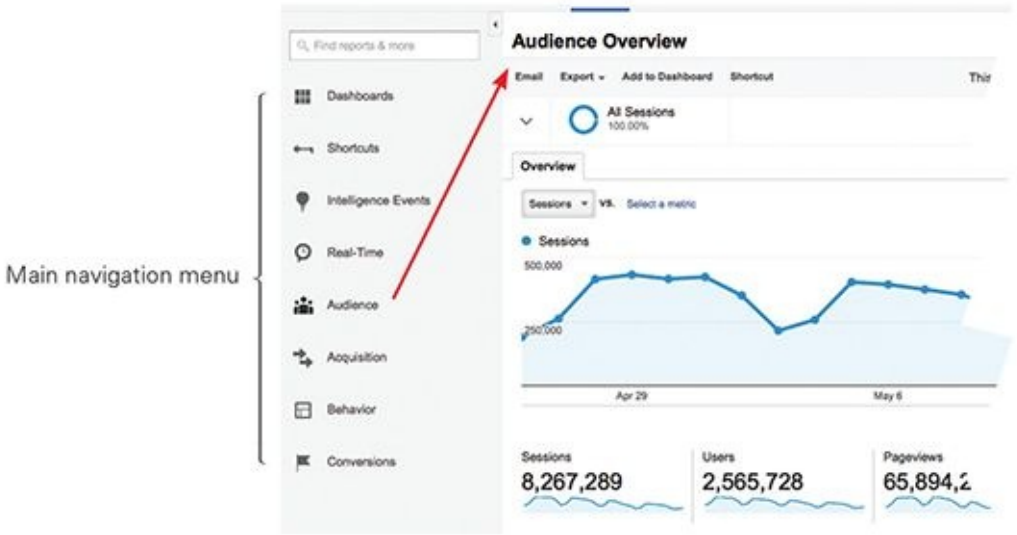


Figure 5.1 *The initial report—Audience Overview—shown when you log in to Google Analytics*

The navigation menu to the left is consistent throughout the interface and is your focal point for finding information. Most of the menu items contain multiple submenus that expand as you click them. Google Analytics has over 100 reports available by default. Don't worry—I am not going through all of these! Chapter 4 contains numerous screenshots if you wish to get a feel for the general report layouts and design.

Custom Reports

On top of the 100+ standard reports described here, you can also build custom reports that combine the data in ways that suit your organization better.

Dashboards "Show me an overview of performance."

A dashboard is a custom report that you build from pieces of other reports. A dashboard lets you view disparate chunks of information, alongside each other. For example, you might want a mini-chart displaying the number of visitor engagements plotted over time, shown alongside a table of your top-performing pages. This will allow you to assess which pages are driving engagements.

The dashboard area is useful for collecting an overview of information from different reports without the need to drill down into the separate reports

themselves. You can create multiple dashboards.

Shortcuts “Take me to that report we examined last week.”

A shortcut is a link, created by you, that takes you directly to a *specific* report configuration. For example, if you have previously investigated a report—perhaps drilled down several layers and applied a detailed segment and a filter—you preserve that specific configuration when you create a shortcut for it. The shortcut link is a helpful one-click access to that report. It means you avoid the need to apply the same settings next time you wish to view the same report.

Intelligence Events “Has anything changed significantly?”

Google Analytics performs some clever predictive analytics by monitoring your data patterns. For example, if it detects that data patterns change significantly, an alert is created in your reports to highlight this for you. Google Analytics determines what constitutes a significant change based on your past performance. Custom alerts can also be created and emailed to you if they trigger.

This is powerful stuff, because rather than analysts having to check each day for significant change in data patterns that can be buried deep in your reports (if traffic to your website drops by 20% compared to this day last week, say, or if product purchases are significantly higher this month than last month) they (and other people) can be alerted automatically—a great second pair of eyes for the digital analytics team that works 24/7 on your behalf.

Real-Time “Show me what’s happening right now.”

As the name suggests, real-time reports are updated in real time— usually within 4 seconds (an amazing engineering feat by Google). These reports are a great way to explore the dynamics of your website and test new campaigns, pages, or website features.

Real-time reports are powerful for tracking anything that is time-sensitive, such as when you unveil a new product, launch a large advertising campaign (at half-time during a football match, perhaps), or have an important PR announcement to make.

To allow for rapid updates, real-time reports are abridged versions of the full reports. That said, they are a comprehensive suite of reports showing traffic volumes; where visitors come from (geo-location); the referral source (the campaigns and other websites providing your traffic); what content is being viewed; in-page events your visitors trigger; and the conversions generated (sales, leads, engagement).

Avoid Real-Time Fixation

Shelby Thayer,[2] a good friend and analytics expert, said, “I have a love–hate relationship with real-time reports. I wish our team were the only ones that knew they existed!”

Her reasoning is that real-time reports are useful for specific things—such as testing and launching a new campaign, page, site, or product. However, beyond this they are eye candy. The issue is that because the charts and numbers update as you watch them, you become fixated. Real-time reports are addictive—a great deal of time can be wasted being mesmerized by the ebb and flow of your traffic. Use them judiciously.

Audience "Who are our visitors?"

The default report loaded when you log in to your Google Analytics account comes from this section. The Audience section is a collection of reports that inform you about your visitors without the use of any personal information: where they are located—accurate to within a 25-mile radius (40 km); their demographics (gender, age group, interest group); whether they are new or returning visitors; how long they stay and the number of pages viewed; visitor frequency and recency (how often and how recently they visit); what devices are used—Windows versus Mac, mobile, tablet, or desktop. Here, mobile and tablet reports can get very specific—right down to make and model of the phone used by your visitors.

Acquisition "Where do our visitors come from?"

Acquisition reports tell you how visitors arrive at your site—which channels and what specific campaigns are driving visitors to you, and how these differentiate when it comes to engagement with your brand and purchasing. For example, if you run a Spring Sale campaign, you can compare its performance across multiple channels. Alternatively, for a given channel, what campaigns are successful?

Campaign information shown in your reports can get very detailed if required. They can show the performance of a particular banner ad (animated versus static); different keywords used by visitors clicking through from your AdWords ads; the success of a specific email sent versus its follow-up—even which link was clicked on within an email; and so forth.

Social interactions are also recorded here—visitors sharing your content via buttons such as Tweet, Like, Pin, or Google+, and conversations that happen away from your website on a particular social network.[3]

Offline Campaign Tracking

With a little know-how, your acquisition reports can also compare the performance of your offline marketing initiatives. There are numerous ways to approach this, with varying degrees of accuracy.

One technique is to use a vanity URL, such as example.com/springsale. This URL is then used in your *offline* advertising only, not published anywhere online (including your own site). In other words, the only way a visitor can be aware of the campaign page is to have seen the vanity URL via your offline campaign.

For *online* marketing efforts of the same campaign, use a separate URL, such as example.com/products/sale/spring. The URL is different and functional—it does not need to be as "attractive" as the URL used for your offline message because online visitors follow links, not the underlying

syntax. Using this method, visitors resulting from online and offline marketing campaigns can be differentiated.

If you adopt the vanity URL method, it is a good idea to also run a paid search campaign in parallel. This is because some visitors will forget the full URL but may remember the product advertised or the buzzwords used. Target these words in your paid search advertising to ensure these visitors also find you.[4]

Behavior “What do our visitors do?”

The behavior report set provides information on what your audience does on your website, such as the content (pages) they view; page load times they experience; whether they use your internal site search tool (and what keywords they use); whether they click on actions within a page (events that are not pageviews); what ads are clicked on; and so forth. Also reported in this section is the performance of any A/B testing you are conducting (displaying different versions of your content to different visitors).

Conversions “What are the important things visitors do?”

Conversion reports, which tell you about visitors who convert into leads or customers, are probably the most important report set for your organi zation. They highlight the metrics that you define as constituting success— the significant engagements that go beyond traffic levels and pageviews. These include detailed reports for e-commerce and engagements (goals). To complement these, Google Analytics includes two sophisticated reports—multi-channel (Figures 5.2 and 5.3) and attribution modeling.

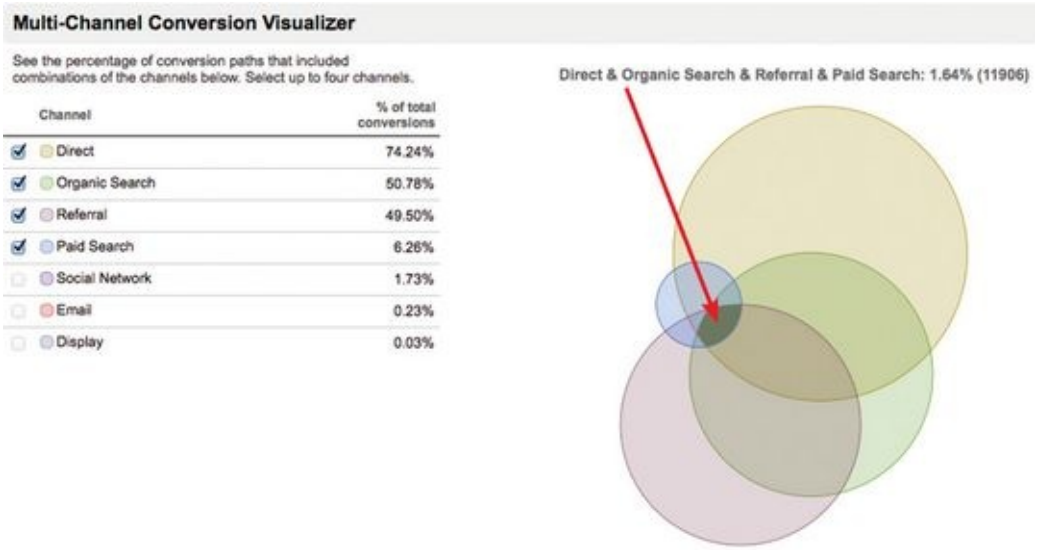


Figure 5.2 *Multi-channel overlap report*

MCF Channel Grouping Path	Conversions ↓	Conversion Value
1. Referral → Direct	10 (1.48%)	$4,000.00 (1.48%)
2. Organic Search → Direct × 2	9 (1.33%)	$3,600.00 (1.33%)
3. Paid Search → Organic Search → Direct	8 (1.18%)	$3,200.00 (1.18%)
4. Direct → Referral	7 (1.04%)	$2,800.00 (1.04%)
5. Organic Search × 3	7 (1.04%)	$2,800.00 (1.04%)
6. Direct × 6	6 (0.89%)	$2,400.00 (0.89%)
7. Direct → Organic Search → Direct	6 (0.89%)	$2,400.00 (0.89%)
8. Referral × 2	6 (0.89%)	$2,400.00 (0.89%)
9. Organic Search → Direct × 18 → Paid Search → Direct × 2	6 (0.89%)	$2,400.00 (0.89%)
10. Organic Search → Referral	6 (0.89%)	$2,400.00 (0.89%)

Figure 5.3 *Multi-channel path report*

Multi-channel reports show what overlaps exist in your marketing campaigns that are driving conversions, such as how many campaigns your visitors touched before they converted, which campaigns were involved, and what order your customers clicked on them.

Attribution modeling lets you experiment with assigning monetary values to each of your overlapping campaigns. Let's illustrate this with a simple example. Suppose a visitor comes to your site on two separate occasions—a click from an organic search engine followed by a click-through from a referral (partner website)—as shown in row 10 of Figure 5.3. During their second visit they convert—either as an e-commerce transaction or as a new lead. How much credit should be given to your organic versus partner marketing efforts? Attribution modeling allows you to experiment with obtaining the best fit.

Attribution modeling is discussed further in Chapter 6 (in ebook 2).

Metrics versus Dimensions

Google Analytics reports consist of two different types of information—metrics and dimensions.[5]

A metric is a number—for example, the number of visitors to your website, the number of conversions from a campaign, the amount of revenue gained, or new leads generated.

A dimension is textual information—for example, the list of your top-performing pages, the most effective campaign names, your best-selling products, and so on.

UNDERSTANDING SEGMENTATION AND ITS IMPORTANCE

Apart from collecting good quality data in the first place (your installation), deciding how to segment your data is the bread and butter of the digital analytics team. Google Analytics has some powerful features that help with this.

Segmentation is the grouping together of similar visitor behaviors to facilitate understanding. Without segmentation, your website reports are a large pot of data containing a mix of visitor types who have different objectives, expectations, and desires for their website visit. As shown in Figure 5.4, visitor types include

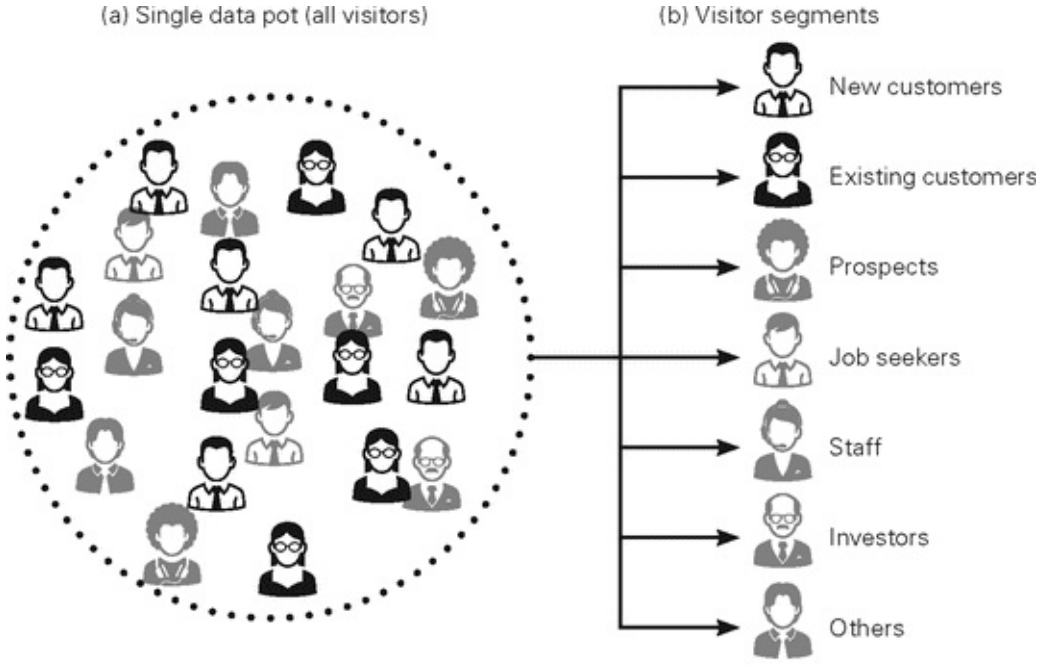


Figure 5.4 *Website visitor types: (a) unsegmented; (b) segmented*

- New customers—people who sign up or purchase for the first time.
- Existing customers—people who need support or who are ready to make a subsequent purchase. They may be potential advocates of your brand, or they may be detractors.
- Prospects—information gatherers who may one day become customers. These can be subdivided by their engagement type—for example, prospects who have subscribed to your newsletter, downloaded your catalog, or submitted a request for information form—or classified as high-, medium-, or low-value prospects. The segmentation techniques for subdividing this group are currently a hot business trend called *marketing automation*.
- Job seekers—from the shop floor to the next potential CEO.
- Staff—people using your website content to facilitate sales and customer service.
- Investors—people who want to know how the business is doing—the vision, the direction, and its execution.

- Others—generally visitors you do not want, such as automated bots (authorized and non-authorized website monitors); spammers; visitors arriving by mistake; or simply visitors of little interest to your organization.

With a single pot of data (Figure 5.4a), very little insight can be gained. Inevitably you will draw conclusions from an average of averages. And when the population is characterized by very different visitor types, the middle ground represents nothing in the real world.

On the other hand, Figure 5.4b represents seven different types of behavior, allowing you to study them separately. At the very least, ensure you can separate your existing customers from your prospects.

Using Segments

Segments do not alter or manipulate your data. They simply group it based on a set of rules you define. You can apply or remove a segment in your reports with one click.

A segment is applicable to all your reports (as per Figure 5.1). As you navigate through Google Analytics, the same segment—and even multiple selected segments—remain in place.

The Power of Segmentation

Consider the needs of your marketing stakeholder who wishes to know how successfully her well-thought-out campaign is creating interest. The purpose is to increase traffic to the website. Therefore, a simple calculation to determine performance during a particular time period is

$$\textit{campaign performance} = \frac{\textit{number of visitors arriving via the campaign}}{\textit{total number of visits}}$$

If there were 1,000 visits via the campaign in an overall visit count of 100,000, the campaign provided only 1% of traffic. That could be construed as a failure considering the investment made, and the campaign might be abandoned.

Now consider Figure 5.5. What if half of the total traffic to the site is from existing customers? And a further half of the remaining visits are from staff, with 10,000 visits from a combination of investors, job seekers, and people not relevant to the business? Now the campaign performance is calculated as a 6.7% improvement in *relevant traffic*—clearly a very different story from what the initial calculation showed. Abandoning this campaign now may be considered a significant lost opportunity. This is the power of segmentation.

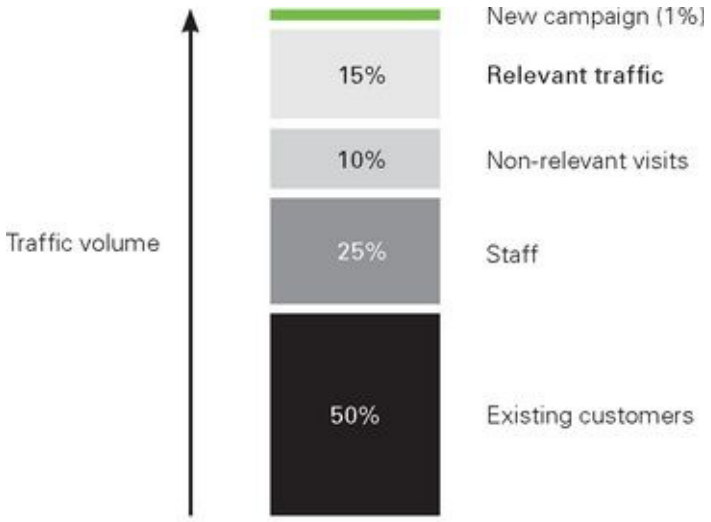


Figure 5.5 *Identifying relevant traffic via segmentation*

☛ *Forgetting to exclude existing customers when calculating campaign performance is a common mistake. I see it often. The reason is usually that the existing customer segment is not tracked effectively*.

Identifying Segment Signals

In an ideal world, the digital analytics team will be able to segment all of the website traffic based on the visitor types shown in Figure 5.4. Doing so requires that website *signals* be available to identify the visitor's segment. Signals are specific actions that are unique to a segment. For example, new customers are those signing up, subscribing, or purchasing for the first time—usually an easy segment to define, as these visitors take clear actions (signals) on your website to identify themselves. Of course, you will need to track these actions.

Beyond the new customer segment, things become harder to classify. For example, if only your existing customers have access to a login area of your website, anyone using that facility can be labeled and segmented as an existing customer. So far, so good, as there is a clear signal (action) for this. However, that does assume existing customers will actually use your login facility in significant numbers. Perhaps the majority prefer to make their purchase at a real-world store, or the value proposition of having a login with you is not strong enough—so existing customers do not create one.

What if no login area exists for current customers?

For these situations, it may be possible to combine other website signals to identify the existing customer segment. Ask the question, "Why would existing customers come back to our website?" If it is to keep informed about product developments, consider setting up a subscription-based newsletter just for existing customers. That way, your customers are informed and won't need to visit your website! However, by embedding "read more" links into your emails, you can track any subsequent click-throughs—solving the identification issue for this segment. Once you have identified it, cookie the segment information of your visitors so that any subsequent visits that do not originate from your subscription emails are also identified correctly.

To facilitate segment identification, the digital analytics team will often need to request changes to the website content, or its architecture, in order to create the signal. Having the web development team on board as a key stakeholder of the analytics project is clearly an advantage.

Similarly, job seekers and investors will view specific areas of your site—assuming such areas exist and are well targeted to your audience. In other words, only people who are job seekers or investors are likely to view that content. Again, the architecture and design of your website may mean that assumption is not valid. To solidify it, discuss potential changes with your web development team so there are clear data signals from these segments. Although job seekers and investors may not be valuable to your sales and marketing stakeholders, it is still important to be able to identify this group of visitors—so that you can exclude them to improve the accuracy of calculations for your other segments (as per Figure 5.5).

☛ *A consequence of adding, or improving, visitor segment signals on your website is improved website usability. Having a much clearer user experience is something that will always stand you in good stead with your visitors*.

Prebuilt and Custom Segments

A number of segment signals in your data are collected by default, such as visitors from mobile devices versus tablets versus desktops (three segments); purchasers versus non-purchasers; people who have visited your site more than once versus single visits; visitors arriving via paid search advertising versus non-paid; visitors who bounce away from your website (single page, no interactions) versus those that stay and browse around. These are all prebuilt segments that you apply to your reports with the click of a mouse—you do not need to identify the segments yourself, as Google Analytics does this for you. The prebuilt segment list is shown in Figure 5.6.

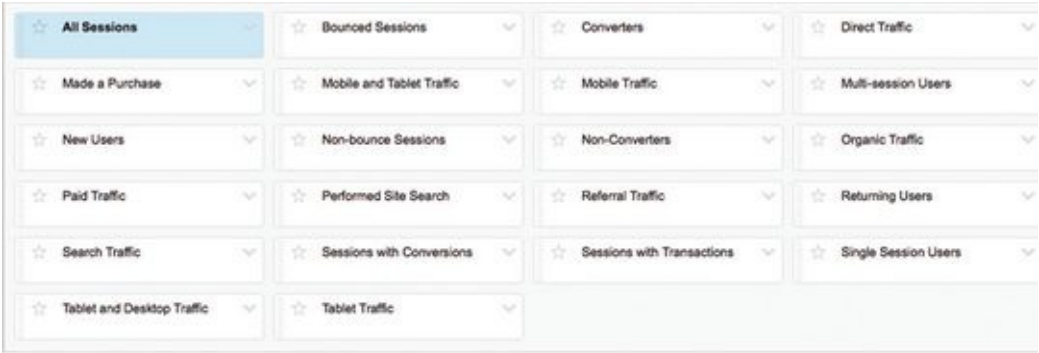


Figure 5.6 *The prebuilt segments of Google Analytics*

Despite the power and ease of use of the built-in segments, things become very interesting when you build custom segments specific to your business. Table 5.1 lists some examples.

Table 5.1 *Examples of Custom Segments*

Segment Name	Type	Description
Remove outliers	Condition match; visit-based	Excludes anyone viewing fewer than three pages, or less than 30 seconds, or more than 10 minutes
Loyal visitors	Condition match; user-based	Includes only visitors who have come to your website on at least three occasions
Big spenders	Condition match; visit-based	Visitors who have purchased with a transaction value of greater than $1,000
High-value customers	Condition match; user-based	Visitors who have purchased a total transaction value of greater than $5,000 over the course of multiple visits
Visits from affiliates	Condition match; visit-based	Includes only referral visitors from affiliates
Cohort subscriptions in May	Condition match; user-based	For visitors who first arrived in May, includes only those who went on to subscribe (in any subsequent visit)
Campaign conversion	Sequence match; user-based	For visitors landing on a specific campaign page, includes only those who went on to become a customer (in any subsequent visit)

The Type column of Table 5.1 has two components:

Condition or sequence match This defines the logic used to match the segment definition. For example, a condition where *medium* = *email* will match all visits that come via email click-throughs. Multiple conditions can be used with AND or OR logic. For example, *medium* = *email* OR *medium* = *organic* will match visitors that arrive from either email click-throughs or organic search engines.

A sequence segment is defined in the same way—using conditions. The difference is that for a sequence segment, conditions must be matched *in order*. For example, where *medium* = *email* followed by *event* = *customer*, a visitor arrives via an email click-through and subsequently becomes a customer. Visitors who arrive via an email click-through and do not become customers are excluded, as are customers who arrived another way. Another example could be when *event* = *watched demo video* followed by *event* = *subscribe*. This would help you understand whether your investment in video production is driving your business forward with subscriptions.

Visit- or user-based If a segment is defined as visit-based, it means the conditions must be matched during the same visit (session). U*ser-based segment* means the same visit or subsequent return visits. This is a very powerful way to define your segments. The caveat is that your users must use the same device and browser combination on their repeat visits for this to work (cookies are device- and browser-specific).

If multi-device usage is common for your website, you can overcome the loss of tracking cookies by implementing the user ID tracking feature of your known visitors.[6]

UNDERSTANDING VISITOR FLOW

Visitor flow is a visualization technique aimed at understanding how your visitors flow around your website. An example is shown in Figure 5.7.

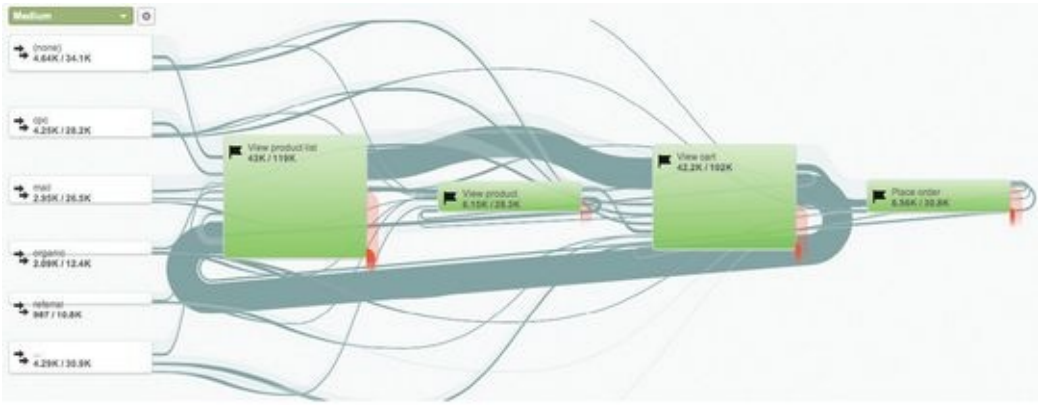


Figure 5.7 *Sample flow visualization report*

You will see this report type in several places—the Audience, Behavior, and Conversions report sections. Essentially, they all attempt to do the same thing—help you understand how visitors move around your website.

What is clever about the flow reports, and the reason I highlight them in this chapter, is that this is *not* path analysis. Google Analytics is not trying to show you every path variation your visitors can take. Few visitors share common paths, so there is no value in analyzing these (assuming analyst time is a premium for you and you value the sanity of your digital analytics team). What Google realized early in its product development is that visitors essentially take random walks through website content until they find what they are looking for. Hence the grouping of *similar* paths is what flow visualizations do.

Google Analytics will automatically group what it considers to be similar paths. However, you can also customize this (referred to as Content Grouping) and apply your own segments—for example, “show the flow of customers only around the sections of my site.” In my view, the data visualization of this technique is industry leading.

☛ *A recent extension to flow visualization reports is the addition of events (non-pageview user actions such as using loan calculator widgets). That is, the same method can be used to visualize page and event flow.*

UNDERSTANDING VALUE

Monetization plays a fundamental role in analysis. Apart from putting a monetary value on what your website is worth—interesting to both transactional and non-transactional websites—if you monetize a process on your website, Google Analytics automatically provides a wealth of additional information.

Monetization allows you to

- Differentiate your most valuable visitors from others. This can be those who purchase the most, comment the most, share the most, download the most, and so forth.
- Identify what content is the most valuable to your visitors. For example, what leads them to become high-value visitors?
- Identify poor-performing pages. These are not necessarily your least popular pages, as conversely pages with low traffic levels may be your most valuable.
- Identify your most valuable campaigns—again, not just what is the most popular at generating traffic; rather, what is most valuable to your business.

There are two ways to monetize your website—using goal values or having a transactional website. The setup of these is discussed in Chapter 3. With your website monetized, reports auto-populate with page and session (visit) values. See Figures 5.8 and 5.9.

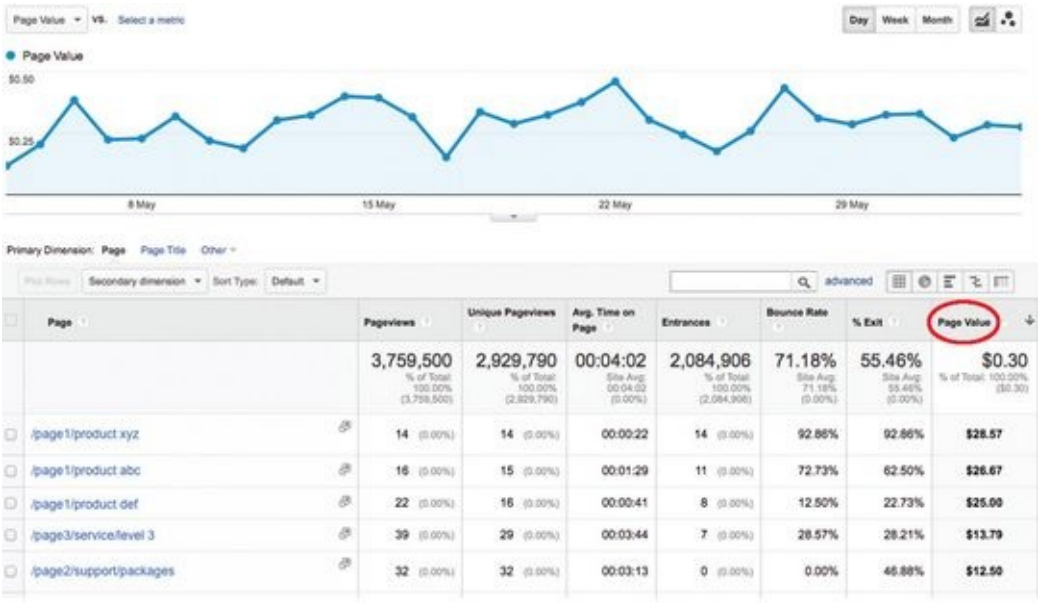
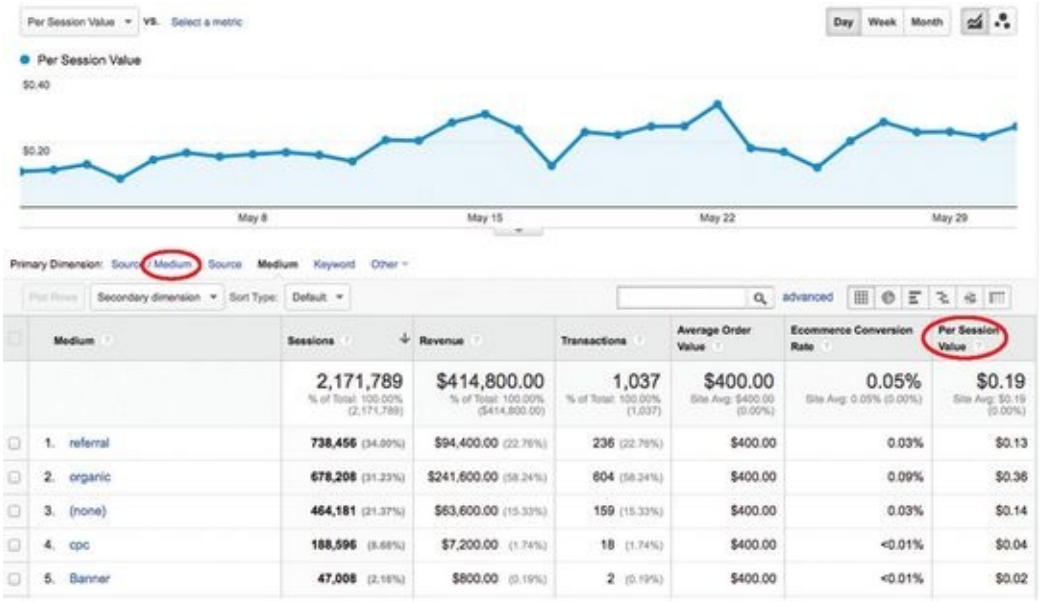


Figure 5.8 *Sample report showing page values*



	Medium	Sessions	Revenue	Transactions	Average Order Value	Ecommerce Conversion Rate	Per Session Value
		2,171,789 % of Total: 100.00% (2,171,789)	$414,800.00 % of Total: 100.00% ($414,800.00)	1,037 % of Total: 100.00% (1,037)	$400.00 Site Avg: $400.00 (0.00%)	0.05% Site Avg: 0.05% (0.00%)	$0.19 Site Avg: $0.19 (0.00%)
1.	referral	738,456 (34.00%)	$94,400.00 (22.76%)	236 (22.76%)	$400.00	0.03%	$0.13
2.	organic	678,208 (31.23%)	$241,600.00 (58.24%)	604 (58.24%)	$400.00	0.09%	$0.36
3.	(none)	464,181 (21.37%)	$63,600.00 (15.33%)	159 (15.33%)	$400.00	0.03%	$0.14
4.	cpc	188,596 (8.68%)	$7,200.00 (1.74%)	18 (1.74%)	$400.00	<0.01%	$0.04
5.	Banner	47,008 (2.16%)	$800.00 (0.19%)	2 (0.19%)	$400.00	<0.01%	$0.02

Figure 5.9 *Sample report showing per session values*

What Does Page Value Mean?

Page value measures the value of a page to your organization in monetary terms —whether you have an e-commerce facility or not. Essentially, it is a way for you to prioritize the importance of your pages. For example, for testing new campaigns or special offers, you will want to start by first looking at pages with the highest page value, as they will have the greatest impact.

To illustrate how this works, consider the following e-commerce example of four visitors who take different paths on your website. A transaction happens on page D and for the simplicity of explanation I fix the revenue at $100:

Visitor 1: page B ► page C ► page B ► **page D**

Visitor 2: page B ► page E ► page B ► **page D**

Visitor 3: page A ► page B ► page C ► page B ► page C ► page E ► page F ► **page D** ► page G

Visitor 4: page B ► page C ► page B ► page F

To calculate the page value of each page visited, Google Analytics allocates $100 to the transaction page (page D) and to each unique page in the path that *precedes* the transaction. Pages viewed after a transaction are ignored. Each page value is then divided by the number of unique times it is viewed by all visitors. The results are shown in Table 5.2.

Table 5.2 *Calculating Page Values*

Page	Revenue / Unique Pageviews	Page Value
A	$100 / 1	$100
B	$300 / 4	$75
C	$200 / 3	$67
D	$300 / 3	$100
E	$200 / 2	$100
F	$100 / 2	$50
G	$0 / 1	$0

Consider page A—this page only occurs once and it happens before a transaction (for visitor 3). Therefore, page A is assigned the revenue from visitor 3. The calculation of its page value is simply 100 / 1 = $100.

Page B occurs eight times in the visitor paths. However, multiple views of the same page in a single path are ignored for the calculation. Hence the number of unique pageviews is 4. As three of these occur before a purchase (for visitors 1, 2, and 3), page B is allocated the revenue three times. The page value of page B is therefore 300 / 4 = $75.

The calculation proceeds in the same way for all the other pages. As you can see, a particular page value is dependent on how many visitors view it on their

way to a transaction *and* the value of the transaction (I use a fixed value of $100 to keep it simple). Page value is therefore a great way to determine the important content on your website.

Because page value is so powerful at highlighting your most valuable pages, I consider it a KPI that can be applied to both transactional and non-transactional websites. For a non-transactional website, your goal values are used for the calculation instead of transaction revenue. See Chapter 9 (in ebook 2) for details on defining KPIs.

> ☛ *Page values are independent of order. Before a conversion, it does not matter if page A comes before page B (or vice versa) for visitor 3*.

What Does Session Value Mean?

Compared to calculating page values, determining the value of a visit (session) is quite straightforward:

$$\textit{per session value} = \frac{\textit{total value}}{\textit{number of sessions}}$$

where *revenue* can be either your e-commerce revenue or goal value. Hence you can view these numbers in either your e-commerce reports or goal conversion reports.

Per session values are interesting when you compare where your visitors come from—as shown in Figure 5.9. Organic visitors are much more valuable than any other channel (at $0.36 per session, nine times more valuable than AdWords visitors). But bear in mind that per session values are based on attributing revenue to the current referral source of the visit. Any previous visits by the same user that did not convert into a purchase are not taken into account.

☛ *A per session value is a good first-guess approximation as to the value of different marketing channels to your organization. A more thorough understanding of value from different traffic sources is obtained from attribution modeling, though this is more complex to set up. Attribution modeling is discussed in Chapter 6 (in ebook 2).*

INTEGRATING WITH OTHER DATA

Google Analytics integrates with a number of other Google products, namely AdWords, AdSense, Google Webmaster Tools, Google Sites, and YouTube. The integrations are usually very straightforward—typically a checkbox and the adding of your Google Analytics account ID—for the data to flow in to your Google Analytics reports.

A particularly strong integration is with AdWords, with a two-way relationship:

- AdWords can import your Google Analytics goal setup. This allows you to use conversion rates (rather than click-through rates) within your AdWords account as the basis for your optimization.
- Google Analytics can import AdWords data (impression, click, cost, and creative information). This allows you to view the complete visitor journey of your AdWords visits within your reports, and measure your return on investment—what money you make from your advertising.

Google Analytics can also integrate with your back office systems, such as your CRM system, using its User ID feature. Essentially, if your visitor authenticates with your internal CRM system that allocates a unique ID value (typically a customer ID), this value can override the anonymous cookie ID Google Analytics uses for visitor tracking.[6]

Implementing user ID tracking means you are tracking real people—as opposed to anonymous cookies used by the default tracking option. These are not quite the same thing, and therefore user ID tracking provides two important benefits to your digital analytics team:

- Improved accuracy—a user ID provides more accuracy than a cookie, as it is unique to the user. It is not degraded by visitors using multiple devices or multiple browsers to access your site (cookies are device- and browser-dependent).
- Multi-device usage—analyze how your visitors use multiple devices (or browsers) to consume your content. For example, research shows 45% of people who research via mobile go on to purchase via desktop or tablet.[7]

The key for taking advantage of User ID is having your visitors identify themselves in some way. Typically this is via a login or authentication facility. In turn, this requires you to encourage your users to create an account in the first instance and for them to actually use it for each visit. For it to work in practice, there must be a real value-added benefit for the user to do so.

Amazon is a great example of the always logged in approach. However, few other organizations have been able to replicate such a strong consumer engagement to be permanently logged in.

☛ *A potential workaround for user login is to have your known visitors click on a customized link within an*

email generated by your CRM system, such as a newsletter where each link contains the user ID for the specific recipient.

IMPORTING DATA INTO GOOGLE ANALYTICS

As I discuss throughout this book, Google Analytics is moving away from being a great visualization and insights tool for web data to being a great visualization and insights tool for *all* your business data. This evolution of the product is called Universal Analytics, and it launched in 2013. A key part of being able to provide insights beyond web data is for organizations to be able to upload their own complementary data sets.

For example, if your internal systems calculate a lifetime value for each customer and keep track of their contract renewal date, you can upload this information into Google Analytics so that it is reported alongside other relevant data. Then, within your Google Analytics reports you can see if there is an increase in web activity from customers who are close to their renewal date and investigate what factors drive an increase in lifetime value.

You can import seven different data types into Google Analytics, as shown in Figure 5.10. All of these are straightforward to implement—you create a comma-separated value (CSV) file with the first row defining the column names, such as Transaction ID and Product Name for refunds. Then upload the file from your computer. The upload is immediate, with file processing taking a little longer, dependent on file size (it may take up to 24 hours for processing).

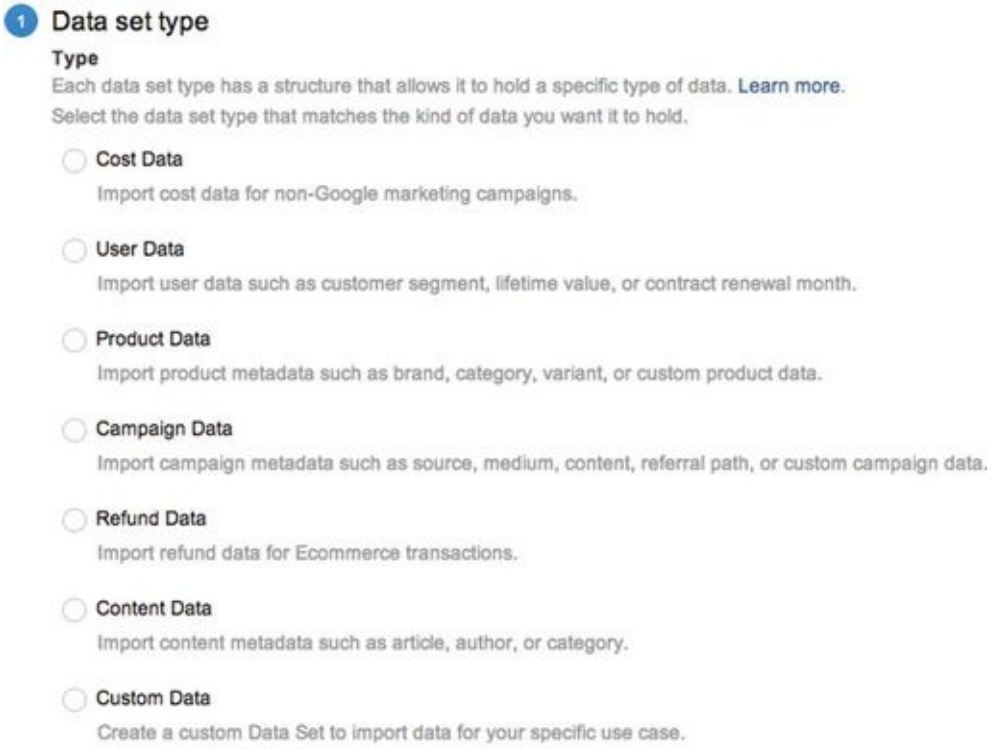


Figure 5.10 *Data import types*

Why might you use data upload?

Clearly there are lots of use cases for this functionality and some are quite obvious from the descriptions in Figure 5.10. Here are three:

Product data upload You have an e-commerce site and your product names change. You want your new product names to be reflected in your historical Google Analytics reports. Uploading a file containing the product SKU (a unique item ID) and new product names allows you to overwrite the old product names in your historical reports.

Similarly, if for some reason it is difficult for you to rename products on your website, you can also use this technique to rename them en masse for future reports—that is, for future transactions.

Campaign data upload Rather than placing your campaign details in your URLs to populate reports with useful campaign information, you can use a simplified code. For example, sending a visitor to *example.com?utm_id=101* labels the visitor as coming from a specific campaign. In order to have human-readable campaign details in your reports, upload a file with the descriptions. For example, for *utm_id=101*, substitute *source = customer_db*; *medium = email*; *campaign name = renewal_message*.

This method not only hides your potentially sensitive campaign information from competitors, but it also greatly simplifies the task of campaign tracking—reducing errors and inconsistencies. It is particularly useful if you have distributed marketing teams.

Refund data upload Processing e-commerce refunds is an obvious need for a business. However, some thought should go into this before applying it to your Google Analytics reports. Take the following example: I purchase shoes from your online store and change my mind when they are delivered (wrong size, wrong color, just not my style). Whatever the reason for my refund, the marketing campaigns that led me to your website and to purchase are still valid. Removing this transaction from your reports will underestimate the value of your marketing efforts. For this scenario, my advice is to keep such transactions in place.

However, there are times when removing a transaction is required, such as when an obvious mistake has been made (wrong item purchased), during site testing, or in the case of fraudulent transactions. To do this, upload a CSV file containing the transaction IDs to be refunded. Partial refunds can also be applied.[8]

What Happens If I Upload a Mistake?

If you find you have made a mistake in your upload file—added a typo, for example—you can upload again to overwrite and correct it. If you later decide you wish to revert to your original data set prior to the upload, simply remove the uploaded file.

☛ *File upload sizes are currently limited to 1 GB per file. If you have a 10 GB product database to upload, split this into 10 files of 1 GB each. There is no limit as to the number of files you can upload*.

SCALING WITH AUTOMATION

The Google Analytics user interface is powerful and intuitive to use. However, it does not scale well when you are managing a large number of websites or users, or when the standard reports require constant customization to meet your needs. This is where automation fits in.

Google has a tradition of providing access to its systems and data via application programming interfaces (APIs). Essentially, if you need non-human access to your Google Analytics reports, the APIs provide the route. APIs enable developers to query Google Analytics and build applications around your data. They do this by means of scripts that automate processes for you so that they can scale. Such scripts can be used, for example, for the management of multiple users or a bespoke query of your data.

Google's approach is to simplify API access and therefore make it available to everyone—though developer knowledge is required. The APIs have proven so successful that an industry of full-blown third-party applications has also evolved to do this for you. If you prefer the ready-made approach, browse the apps by Google Analytics Technology Providers at www.google.com/analytics/apps/search/apps.

If you prefer a more hands-on approach and have the resources available (a web developer on hand), there are numerous specific APIs available.[9] These fall into four groups:

Reporting APIs These allow you to query any part of your collected data, for any date range, including your real-time reports. Data can be accessed and saved into a spreadsheet or database, for example. This method is extremely powerful for building your own dashboard, or when you wish to combine Google Analytics data with other data sources for further manipulation and display—such as your sales data, call center data, marketing plan, or social media activity. Most third-party tools for Google Analytics use the reporting APIs.

Provisioning APIs These allow you to create new accounts, properties, views (report sets), and new users. This is very useful if you are an Internet service provider or hosting company, where you may have thousands of accounts and users to set up—a laborious task if you were to use the user interface. An example is automatically setting up and pre-configuring a Google Analytics account when a new customer signs up.

Management APIs This is a collection of APIs allowing the automatic management of your Google Analytics configuration. Create or edit data filters en masse; change user permissions; bulk connect AdWords accounts, and so forth. Management APIs are aimed at large organizations with multiple websites, multiple Google Analytics accounts, or a large number of users requiring data access.

Embed API This is an API to create and embed a dashboard of your Google Analytics data on a third-party website—this could be your intranet site, for example. The Embed API allows you to build custom Google Analytics visualizations and layouts to meet your stakeholder needs.

UNDERSTANDING REPORT SAMPLING

Sampling is a standard statistical approach when dealing with large volumes of data. In business and medical research, sampling is widely used for gathering information about a large population—for example, clinical trials of new drugs, predicting the result of an election, or manufacturing quality control. These all rely on the use of random samples as a method to estimate characteristics of the whole population.

The vast majority of Google Analytics reports you are likely to view are *not* sampled—regardless of data quantity. These are the 100+ default reports that you have access to when you log in. They are precalculated aggregates and always contain the complete data set. However, if you apply segments, add secondary dimensions, or build your own custom reports, Google Analytics has to go back and reprocess your raw data tables as an ad hoc request. This is when sampling can occur, applied in order to make efficient use of Google's processing capacity.

The purpose of sampling is to ensure your reports are loaded quickly and you are not waiting for long periods of time while the data is reprocessed. The sampling thresholds are shown in Table 5.3 for both Google Analytics free and Premium (paid-for).

Table 5.3 *Google Analytics Sampling Thresholds*

	Google Analytics Free	Google Analytics Premium
Sampling threshold (visits)	500,000	10 million*

* This number is constantly increasing as Google allocates more resources to Premium processing.

When the thresholds of Table 5.3 are reached, Google Analytics takes a random sample of your data at the visit level (only whole visits included[10]), uses this sample to perform its calculations, then scales the numbers back up to represent the total in the created report. For example, if the total pot of data in your custom report includes 5 million visits, a random sample rate of 10% is used for the sample calculations, which are subsequently multiplied by ten in the displayed report to estimate the totals. If you were using Premium, this report would not trigger sampling.

Data Collection Is Never Sampled

Google Analytics does not sample your data at the point of collection or processing. The only exception to that rule is if you break the terms of service (ToS). This states you can use Google Analytics for free for up to 10 million data hits per month (approximately 1 million visits per month). If your site regularly receives more than that, you need to upgrade to the paid-for Premium product, or throttle the data flow you send to Google—that is, sample your own data collection. If you fail to do that, Google reserves the right to do that for you.

☛ *As a rule of thumb, if your site receives less than 100,000 visits per month, it is unlikely you will come across report sampling*.

To indicate that a report is generated from sampled data, Google Analytics displays a notification box at the top right of the report screen, as shown in Figure 5.11. Although it doesn't say so explicitly, this means sampling has taken place (20.72% of sessions sampled).

Figure 5.11 *Report sampling notification*

Pitfalls of Sampling

Sampling is a great way to obtain estimates of characteristics for the whole population. However, if you are looking for estimates of a much smaller subpopulation of the total, error bars can become very large. It is the needle in a haystack problem. With only a handful of needles in the haystack, there is a strong possibility of taking a sample of hay containing no needles at all (or very few). When Google Analytics scales the sampled number back up, the totals arrived at for the estimated count of needles can be way off—even meaningless.

Here's a real-world example:

Your conversion rate on average is 1% for global visitors, but for your particular analysis you are interested only in US visitors from Arizona—this accounts for approximately 1% of all traffic. You apply a segment to obtain this for the month of May. As you have a total of 5 million visits in your data set, sampling is invoked at a sample rate of 10%.

For the sampled size of 500,000 visits, you are looking for only 50 conversions (1% of 1%). Importantly, that comes with an assumption that all conversions are equally distributed throughout your data. However, maybe most of your conversions happened on one particular day— random sampling cannot take this into account. Nor can it take account of a smaller fraction of users making multiple conversions. Hence, it is entirely feasible for the sample taken to only include 25 conversions. That is 50% less than the true value.

Avoiding Sampling

Sampling can be avoided by analyzing multiple smaller sets of data that are not sampled, then summing these metrics to get the totals. For example, rather than using data for the whole month of May, split it into 31 single days to avoid the sampling thresholds, then sum. This is a perfectly valid technique for certain metrics, and there are a number of API tools that provide this feature.[9]

However, be aware that summing in this way only works for visit (session)-based metrics and dimensions, such as number of visits, time on site, bounce rate, number of pageviews, goal completions, transactions, referral sources, device used, and so forth. It does not work for any metric or dimension that is user-based, metrics such as number of unique visitors; demographic information (age, gender); behavior information (new versus returning visitors, frequency and recency metrics); multi-device usage; or attribution modeling. The important point is that users are not additive—sessions are.

Consider report sampling as a sign of success—that is, you are receiving so many qualified visitors that reporting all their activity becomes a time-consuming and resource-intensive process. If you discover that sampling is a constant thorn in your analysis, consider investing in the paid-for Premium version of Google Analytics (discussed in Chapter 2).

☛*The API automation required to avoid sampling means the extraction of data outside of Google Analytics. Report visualizations such as visitor flow and multi-channel funnels are not available for the extracted data set*.

SPEAKING OF FEATURES …

When I hear this …	I reply with …
We need to compare reports from 15 regional web properties side by side. What's the best approach for this?	This is something Google Analytics cannot provide for within its user interface. Therefore, build your own dashboard to do this with the Embed API,[11] or use a third-party tool with Google Analytics connectivity built in.[12] The latter gives you the added benefit of combining with other non–Google Analytics data sources.
We wish to integrate Google Analytics with our CRM and set the unique reference key as the username—that is, match up usernames in our CRM with usernames in Google Analytics. Can this be done?	Yes and no. The integration is entirely possible using theUser ID feature of Google Analytics.[6] However, this must be done in an anonymous way—such as using your CRM customer ID to overwrite the Google Analytics cookie ID. No personal information can be sent to Google Analytics, as this is a violation of the ToS.
We have a sampling issue but cannot afford the upgrade to Premium (yet!). Is there a legitimate alternative?	It is perfectly acceptable to split your data into small subsets to avoid the sampling thresholds and then sum these to obtain the results for unsampled data. This is an API technique you can build yourself, or there are third-party tools available.[9]

CHAPTER 5 REFERENCES

1 The term *Internet of things* was proposed by Kevin Ashton in 1999: http://en.wikipedia.org/wiki/Internet_of_Things.

2 Shelby Thayer, Director of Web Strategy, User Experience, and CRM. Penn State University: www.linkedin.com/in/shelbythayer

3 Conversations that happen away from your website are shown within Google Analytics if the social network used by your visitors is participating in Google Analytics' social data hub initiative. Unfortunately, that does not include Facebook or Twitter: https://developers.google.com/analytics/devguides/socialdata/.

4 I have written a detailed white paper specifically about tracking offline marketing. Download from http://brianclifton.com/tracking-offline.

5 Understanding metrics and dimensions: https://support.google.com/analytics/answer/1033861

6 Tracking your authenticated users anonymously with a unique user ID is described at https://developers.google.com/analytics/devguides/collection/analyticsjs/user-id.

7 Custom Nielsen study, commissioned by Google (2013)—Mobile Path to Purchase: www.thinkwithgoogle.com/research-studies/mobile-path-to-purchase-5-key-findings.html

8 Importing refund data is covered at https://support.google.com/analytics/answer/6014861?ref_topic=6015090.

9 The full list of available APIs is found at https://developers.google.com/analytics/devguides/reporting.

10 Google describes report sampling in detail at http://code.google.com/apis/analytics/docs/concepts/gaConceptsSampling.html

11 The Embed API allows you to build your own dashboard with Google Analytics data: https://developers.google.com/analytics/devguides/reporting/embed/v1/.

12 Tableau (www.tableausoftware.com) and Klipfolio (www.klipfolio.com) are two well-respected data dashboard tools that integrate with Google Analytics as well as other data sources.

Appendix: Terminology Explained

The definitions presented here are specifically in relation to their usage with Google Analytics and digital analytics in general.

A

A/B split testing

The science of experimenting with the display of different versions of content to different sets of visitors—one version (A) shown to one set of visitors and another version (B) shown to others. The purpose is to statistically determine which version is favored over another by your *customers* and *potential customers*, rather than guessing.

Version preference is determined by measuring the difference in subsequent goal conversions. The version that drives more visitors to a call to action—such as a click on an “add to cart,” “contact us,” or “read more” button—is the preferred version. This is a statistical method that utilizes sampling techniques (Which visitors see which version?) and significance tests (Is the measured preference real or by chance?). It can be applied to content on your website, app, or advertising campaign.

Although the technique is often called A/B testing (or just split testing), more than two versions (A/B/C/D …) can be tested in the experiment.

AdWords, AdSense, and DoubleClick

These are advertising platforms from Google.

AdWords is an auction system for text ads where advertisers bid to have their ad displayed based on a user’s search term. It is a type of *contextual* advertising. The advertiser only pays if their ad is clicked on—there is no charge for display. Hence, the format is also often referred to as pay-per-click or cost-per-click advertising. AdWords ads are served alongside organic Google search results and other Google properties, such as YouTube and Google Maps. They can also be displayed on Google partner sites such as AOL and other portals.

AdSense is a tool for Google partners (for example, portal sites) to manage how Google AdWords ads are displayed on their site. Partner sites that display AdWords ads receive a share of the revenue generated by any visitor click-throughs.

DoubleClick is a tool for advertisers to manage display (banner) advertising. As with AdWords, DoubleClick uses the technique of contextual advertising for displaying banners relevant to a visitor’s query or preference.

API

Abbreviation for *application programming interface*. An API is a standardized protocol to share data between applications—for example, between Google Analytics and an Excel worksheet. Google is a stalwart of the API method and virtually all of its products provide one on a free-to-use basis (within limits). For Google Analytics there are developer libraries available for the end user to experiment with API calls, and there are third-party vendors with prebuilt

applications to achieve certain tasks.

A common use of Google Analytics APIs is the extraction of data into a simplified dashboard format.

attribution modeling

Attribution is the term used to describe who gets the credit for a conversion. That is, for a visitor that comes to your website multiple times via different referral sources, which referrer should get the credit when that visitor becomes a customer? Historically, the last referral source has been assigned the credit, as it was the easiest to determine.

Attribution modeling allows the credit to be *shared* among some or all of the referral sources. For a visitor who converts on their third visit, you can attribute 33% to each referral, split as 40%, 20%, 40%, or use some other model. Attribution is discussed in detail in Chapter 6.

B

bounce rate

The percentage of visitors that bounce away from your site—that is, view only a single page with no other *in-page* interaction.

Generally, a bounced visit is considered a poor user experience, and therefore the bounce rate is often a KPI that organizations wish to minimize. The hypothesis is that a visitor, even with a low interest level, would at least click onto a second page or interact with something within a page, such as a loan calculator widget.

The reasoning is valid so long as you do not deliberately design your site to cater for a bounced experience. Even if you create the perfect one-page article, your visitors will likely wish to tell you this (rate or comment on it), read related articles, connect with the author, share it with others, subscribe, click a related ad, and so forth. Any such action results in a visitor’s not bouncing.

C

campaign tracking

Required to track specific marketing campaigns. The method consists of appending parameters to landing page URLs. This can become very precise—for example, a visitor click-through from a text link in an email campaign that is sent to your California customers only.

Up to five campaign parameters can be specified: *source* (where you place your ad); *medium* (the digital channel used); *campaign* (the campaign name); *content* (type or version of content used); *term* (keywords targeted) if the ad is related to search marketing.

These parameters are not required for AdWords if your account is linked to Google Analytics. Campaign tracking is described in detail in Chapter 6.

channel

A high-level grouping of campaign information. For example, the mediums “cpc,” “ppc,” and “paidsearch” are all classified into the “Paid Search” channel by Google Analytics.

Often the terms *channel* and *medium* are used interchangeably. However, they are subtly different within Google Analytics reports. See also *campaign tracking* and *medium*.

conversion (synonymous with goal)

Any action that is considered a successful visitor engagement. These are pageview or event specific. Examples: a completed transaction (becoming a customer); a submitted contact request form (becoming a lead); the sharing of your content on a social network (becoming an brand advocate).

Up to 20 goals can be defined for each report set in Google Analytics, and each can be monetized by adding a goal value. The *conversion rate* is the number of completed goals divided by the number of visits, expressed as a percentage.

cookie

All the main web analytics tools use cookies as the basis for their tracking methodology. A cookie is a small text file that a web server transmits to a web browser so that it can keep track of the user’s activity on a specific website. The visitor’s browser stores the cookie information on the local hard drive as name–value pairs. The latest version of Google Analytics uses a single cookie to anonymously identify a visitor—using an anonymous visitor ID.

cost-per-click advertising (CPC)

A method of paying for advertising based on the number of click-throughs received (not the display of the ad). Almost all search engines use an auction bidding system to sell advertising alongside organic search results.

Also referred to as a pay-per-click (PPC) ad model, this technique is common for selling advertising space on sites in general—for example, news portals. Google dominates the market with its systems AdWords, AdSense, and DoubleClick, though the originator of the method was GoTo.com (acquired by Yahoo! in 2003). See also *AdWords, AdSense, and DoubleClick*.

cross-domain tracking

The tracking of a visitor that traverses your subdomains or third-party domains, such as the following path:

www.example.com ⇨ shop.example.com ⇨ www.paymentsite.com ⇨ thankyou.example.com

If cross-domain tracking is not correctly set up, you can double-count your visits and lose valuable referral information. This is because the tracking method employed (cookies) is specific to the exact hostname of the website setting them.

custom dimensions and metrics (*see also* visitor labeling)

All Google Analytics reports consist of showing dimensions. By default, there are over 200 available. You can extend this list by defining your own—typically by uploading data to populate your reports, or applying a label to differentiate and segment your visitors. Examples include defining a customer lifetime value; adding author and subject categories for publishing sites; and labeling your high-value customers.

The free version of Google Analytics allows you to define 20 custom dimensions and 20 custom metrics. For Premium, these limits are 200 and 200, respectively.

D

dimensions and metrics

Google Analytics reports display two types of information—metrics and dimensions. A metric is a number—for example, the number of visitors to your website, the number of conversions from a campaign, the amount of revenue gained, or the number of new leads generated. A dimension is textual information—for example, the countries your visitors come from, the list of top-performing pages, the most effective campaign names, your best selling products, and so on.

See also *custom dimensions and metrics*.

E

event tracking

Events are *in-page* visitor actions, as opposed to pageviews—for example, interacting with a loan calculator, scrolling through an image carousel, or zooming or panning around a map. Events are not tracked by default in Google Analytics and therefore must be defined and set up by you. An event can be classified with a category, action, label, and value.

F

funnel visualization

Directly analogous to a traditional sales funnel. In Google Analytics you can visualize the steps visitors take in order to convert (become a new lead or customer). For example, you can examine the drop-off rate of a shopping cart purchase.

G

GATC

The Google Analytics tracking code (GATC) is the snippet of code that is required to track visitors and users. For web pages, this is JavaScript code. Mobile apps use their own programming environment (Objective-C, Java, and so forth).

goal (*see* conversion)

K

key performance indicator (KPI)

Key performance indicators are a set of metrics (defined by senior leadership) to measure progress toward organizational goals. For digital analytics, I define a KPI as *a measurement used to evaluate success*.

This is similar to defining a goal for Google Analytics reporting. The difference is that a goal is the language of your analysts. KPIs are the language of your business. KPIs are the success metrics for the business overall and often combine data that is outside of the Google Analytics realm, such as staffing levels or call center volume.

L

landing page

The first page a visitor views during their visit. Also referred to as the entrance page.

M

medium

When you define a marketing campaign, you can specify up to five parameters for Google Analytics reporting (see *campaign tracking*). The *medium* parameter refers to the category, or type, of content that contains a link to your site, such as “email,” “paid search,” or “organic search.”

Medium is a precise value that may be set by the Google Analytics tracking code or defined explicitly by the user. It should not be confused with *channel*, which is a higher-level grouping of campaigns. Campaign tracking is discussed in Chapter 6.

O

organic search engine

The main search engines (Google, Bing, Yahoo, and so forth) return two types of results for a visitor's search query—organic results and paid results. Paid results are just that, advertisements (see *cost-per-click advertising*). Organic results, on the other hand, are earned results. That is, you cannot pay for placement; they are provided free.

Your web page ranking in the organic results—whether your page appears at the number 1, 10, or 1,000 position—is determined by multiple factors. The key ones are *relevancy*—does your page contain content relevant to the search term used?—and *authority*—how authoritative is your page compared to the plethora of other web pages of similar content (often determined by seeing how many other web sites link to your page or mention your content)?

My description is a huge simplification of a complex subject in computer science, one that employs the latest artificial intelligence techniques so that a computer network, such as Google, can emulate what a human's intent is when they type in a search query. An industry has grown up around improving a site's search engine visibility, known as search engine optimization (SEO).

outbound link

A link, placed on your website, that points to a third-party domain. The third-party domain can be another web property you own or completely separate, such as a reseller or partner website. A visitor who clicks on an outbound link leaves your website. Hence, outbound links are not tracked in Google Analytics by default—you need to make a modification to your tracking code.

P

pageview (*see also* virtual pageview)

This is the default Google Analytics tracking method (hit type) for websites and is the fundamental measurement dimension. Reports also show *unique pageviews*. This aggregates pageviews that are generated by the same user during the same session. A unique pageview therefore represents the number of sessions during which that page was viewed one or more times.

Note that if a visitor returns to the same page URL, but the page title has subsequently changed, this is treated as a new unique pageview. That is, the combination of URL and page title defines if a page is unique in your Google Analytics reports.

R

referrer and referral

A referrer is a source of traffic to your website. Generally, this means any traffic source that is not a direct visit (where the visitor typed in your web address directly—either from memory or using a browser bookmark). A referrer can be from a specific campaign email, organic search engine, advertising campaign, another website, app link, and so forth.

However, Google Analytics automatically classifies referrer visits from organic search engines and AdWords (if linked). Therefore, the term *referral* is used specifically to identify those visits that have come via *another* website linking to you.

For Google Analytics to work effectively at identifying your traffic sources and referrers, campaign tracking must be set up. Campaign tracking is discussed in Chapter 6.

S

site search

The term used to describe your internal search engine, if you have such a facility on your website.

STAG

The acronym I use for the site tracking assessment and guidelines document. It is an installation document that guides the web development team on what needs to be done in order to provide a best-practice setup (a data quality score of above 50). It sets out the required tracking methodology and the reasoning for the requirement. Building a STAG document is describe in Chapter 3.

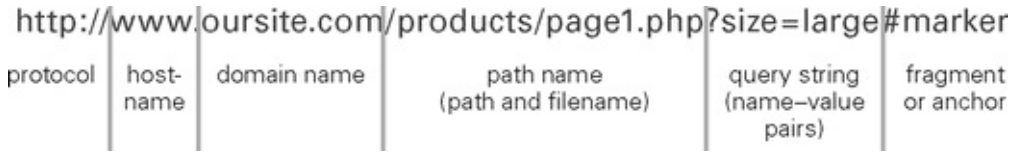
U

Universal Analytics

The new name for the Google Analytics tracking code (as of May 2014). The term refers to the fact that Google Analytics is no longer simply a web measurement tool; it can be used to collect data from any Internet-connected device, such as mobile apps, barcode scanners, event turnstiles, parking lots, door sensors, and so forth. The potential for Universal Analytics is discussed in Chapter 6.

URL

Uniform resource locator—this is the address as written in your browser address bar. It consists of multiple parts. The default tracking behavior for Google Analytics is to report on everything after the domain name, with the exception of the fragment (anchor). The domain can also be included in reports with the use of an include filter.



V

virtual pageview

The substitution of an alternative pageview URL for the default URL captured. Virtual pageviews can be used if a real URL is not generated by a visitor's behavior, or to construct more meaningful URLs than would otherwise be reported.

A typical example is when submitting a form, such as a subscription request. If the form URL does not change when a visitor submits their information, Google Analytics cannot differentiate between a form view and a form submission. This is overcome by sending a virtual pageview on the form submission.

visitor labeling (*see also* custom dimensions and metrics)

This is the application of a label used as a form of classification. The purpose is to allow you to segment your reports based on your own classification names. For example, a label named Visitor Type can have the values Customer, Prospect, Staff, Student, and so forth.

Visitor labels can be applied at four levels:

- Visitor—the label is associated with the visitor on the current and any return visits (applied via a cookie). For example, Visitor Type = "customer."
- Session—the label is associated with all data hits of the visit for which it has been set. For example, Logged-In = "yes."
- Hit—the label is associated with the single hit (a page or an event) for which it has been set. For example, Page Interaction = "none."
- Product—the label is associated with the product for which it has been set (Enhanced Ecommerce only). For example, Product Meta = "discounted item."

Up to 20 different labels can be set in the free version of Google Analytics (200 for Premium). Each label can have unlimited names. The method to apply labels is via setting a custom dimension.

Index

Note: This index covers topics in this ebook only (ebook 1). Ebook 2 has its own index. Many topics are discussed in both volumes.

About the Author

Brian Clifton (PhD) is an analytics advisor, author, and educator who specializes in performance optimization using Google Analytics. He is recognized internationally as a Google Analytics expert. His book *Advanced Web Metrics with Google Analytics* was issued in three editions and has sold more than 70,000 copies. It is used by students and professionals worldwide.

As Google's first head of web analytics for Europe (2005–2008), Brian built a pan-European team of specialists whose legacy was the online learning center for the Google Analytics Individual Qualification (GAIQ).

Brian is a guest lecturer at University College London and at the Stockholm School of Economics. You can hear him speak at numerous conferences around the world—particularly in Europe, where he presents on web measurement, digital marketing, content optimization, and how these can all interlink to create a successful online business strategy. Brian is currently director of data innovation at Search Integration AB.

Successful Analytics: Gain Business Insights by Managing Google Analytics
By Brian Clifton

Ebook 1, chapters 1–5:
ISBN 978-1-910591-01-7

Also available

Ebook 2, chapers 6–10:
ISBN 978-1-910591-02-4

Trade paperback:
ISBN 978-1-910591-00-0

Published by
Advanced Web Metrics Ltd
1st Floor Premier House
46 Victoria Road
Burgess Hill, West Sussex RH15 9LR
United Kingdom
www.advanced-web-metrics.com

See www.brianclifton.com for resources related to the book or to contact the author.

Cover: Mattias Lager, www.addkolon.se
Chapter opening illustrations: Cris Hammond, www.crishammond.com
Design and composition: Dick Margulis, www.dmargulis.com
Proofreading: Katharine Wiencke, kwiencke@verizon.net
Index: Marilyn Augst, www.prairiemoonindexing.com