

Assessing Your Data Quality



Auditing your Google Analytics Data

Foreword

This document is chapter 4 from the book: “Assessing Your Data Quality”. In non-technical terms it describes the entire manual process of building a data quality audit of your Google Analytics setup.

A little background...

In 2015 I published [Successful Analytics: Gain Business Insights by Managing Google Analytics](#). It was my fourth book about Google Analytics and data quality. A major part of the work was to describe my audit method – that is, how to quantify your data quality and simplify its visualisation. The method uses a weighted scorecard approach for the visualisation, and builds a compound metric, the Quality Index, as a measure of how close a setup is to a best practice implementation.

As I hope you will discover, measuring your data quality is a game changer – you will realise just what you can do (and more likely cannot do) with your current data. That said, manual audits are a laborious process. Wouldn't it be a lot more fun easier to automate it with some smart technology? And that is how the idea of [Verified Data](#) was born.

Thank you for downloading this document. If you found it useful please share it with others and connect with me on LinkedIn.

Brian Clifton



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.

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

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 month¹—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.

- 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.
 - 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.
-

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.

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

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	Yellow	5
2	Tracking code deployment	1.0	Green	10
3	AdWords data import	1.0	Red	0
4	Site search tracking	1.0	Red	0
5	File download tracking	1.0	Red	0
6	Outbound link tracking	1.0	Red	0
7	Form completion tracking	1.0	Red	0
8	Video tracking	N/A	–	–
9	Error page tracking	0.5	Red	0
10	Transaction tracking	2.0	Red	0
11	Event tracking (non-pageviews)	1.0	Red	0
12	Goal setup	1.0	Red	0
13	Funnel setup	1.0	Red	0
14	Visitor labeling	1.0	Red	0
15	Campaign tracking	1.0	Yellow	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.

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.

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.




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 \times 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*.

Page 1 of the audit report (the scorecard summary) is the most important part of the report—as a manager, your focus is on this page. The

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

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.

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>.*

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 file-name 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):

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

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

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

Item 2, Tracking code deployment: The weighted score is $1.0 \times 10 = 10$.

Item 3, AdWords data import: The weighted score is $1.0 \times 0 = 0$.

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

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

Again, this is straightforward. The *sum of weights* \times *status value for green* is the sum of weights multiplied by 10. Using the example in Figure 4.1, this is $14.5 \times 10 = 145.0$. 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?

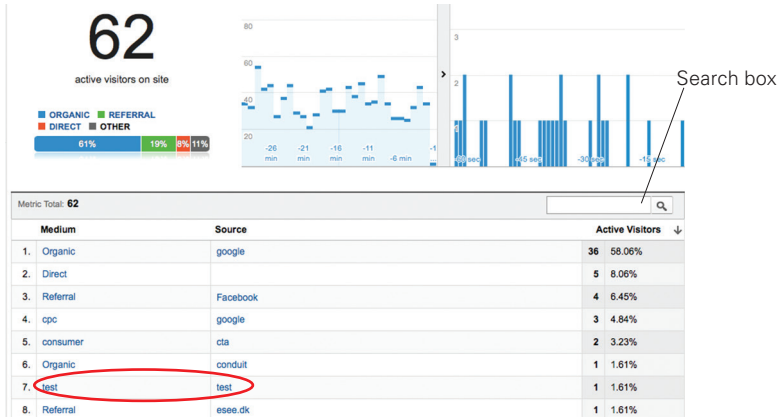
Regardless of how a visitor arrives at your website, as long as you have deployed your tracking code, Google Analytics will track by default the visit, the visitor (the same visitor may make several visits), and the pages viewed by your visitors. That is a simple yet powerful data set that allows more 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 visitors 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.

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.

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:

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.

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.

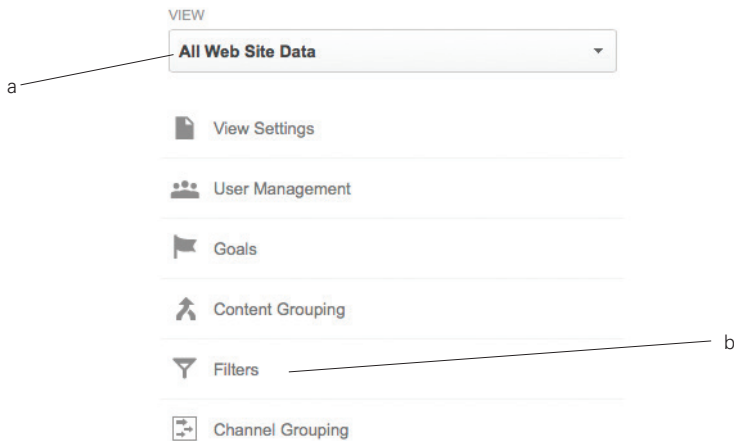


Figure 4.2 The Google Analytics Administration area for a report set: (a) assessing different report sets (views); (b) assessing applied filters

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.html³).

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

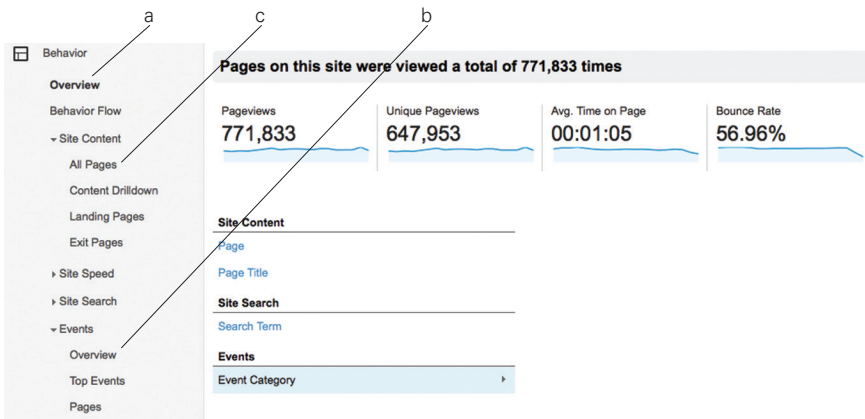


Figure 4.3 The Google Behavior report area: (a) pageviews; (b) events; (c) page content

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.

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 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

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

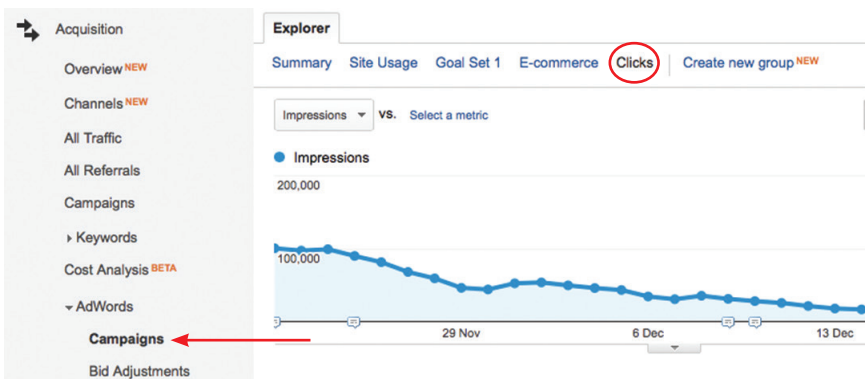


Figure 4.4 Checking AdWords data

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.

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.

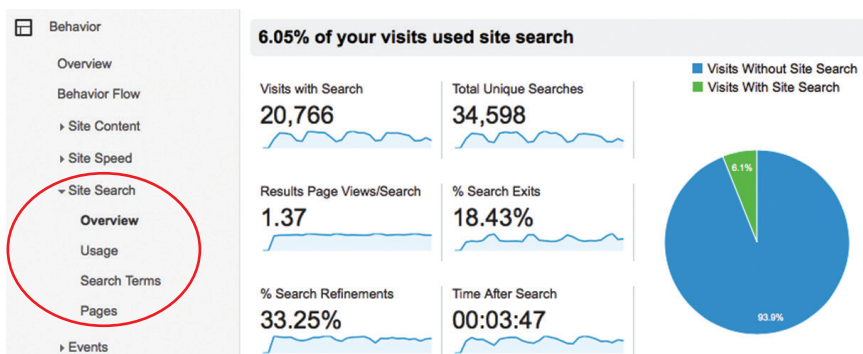


Figure 4.5 Checking your site search data

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?

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.

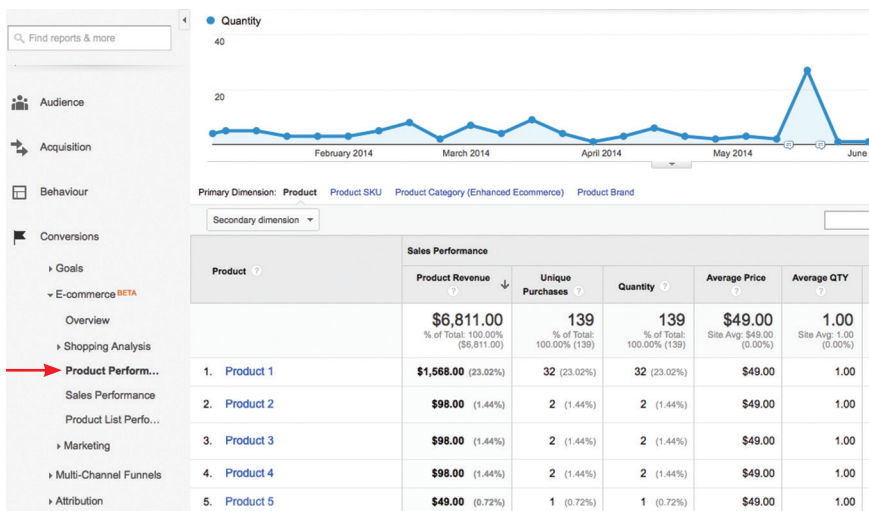


Figure 4.6 Example of a transaction report

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).

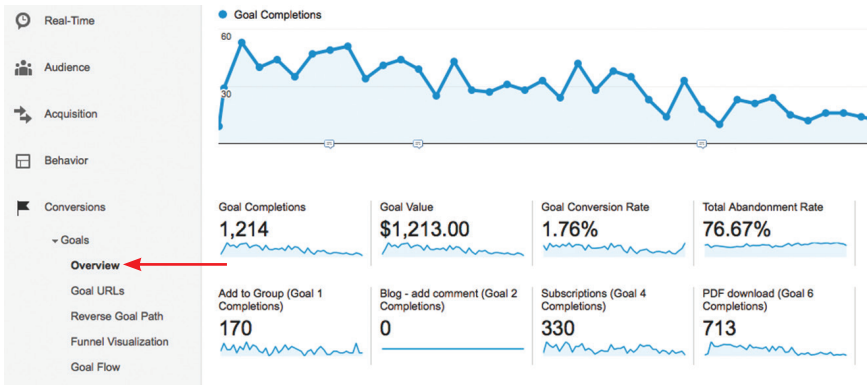


Figure 4.7 Sample goal conversion overview report

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.

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.*

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.

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*).

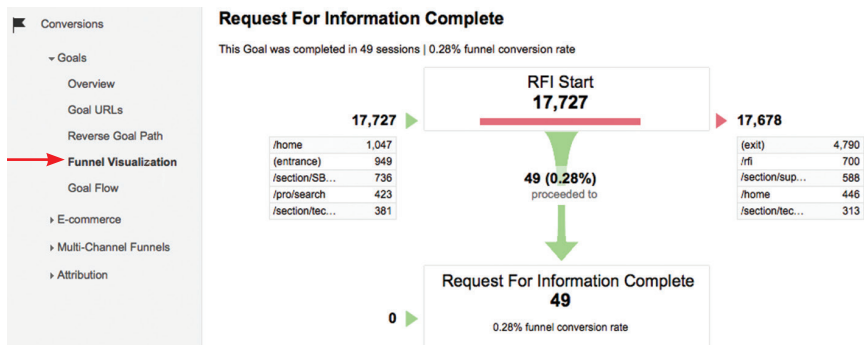


Figure 4.8 Example of a funnel visualization report

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.”

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.

The screenshot displays the Google Analytics Admin interface. On the left sidebar, under 'Custom Definitions', 'Custom Dimensions' is highlighted with a red arrow. The main content area shows a 'New Custom Dimension' configuration page with a quota of 1/20. Below this is a table listing the configured custom dimensions:

Custom Dimension Name	Index	Scope	Last Changed	State
1. AVP remembered	1	User	13 Dec 2013	Active

At the bottom of the table, it indicates 'Show rows: 10', 'Go to: 1', and '1 - 1 of 1'.

Figure 4.9 Viewing visitor labels for a Universal Analytics deployment

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.

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.

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.

Acquisition		Acquisition		Behaviour	
Medium		Visits	% New Visits	New Visits	Bounce Rate
		30,592 <small>% of Total: 100.00% (30,592)</small>	43.60% <small>Site Avg: 43.58% (0.05%)</small>	13,338 <small>% of Total: 100.05% (13,331)</small>	19.06% <small>Site Avg: 19.05% (0.00%)</small>
<input type="checkbox"/>	1. (none)	13,169	38.43%	5,061	11.40%
<input type="checkbox"/>	2. organic	8,574	42.76%	3,666	18.28%
<input type="checkbox"/>	3. cpc	4,515	45.98%	2,076	40.42%
<input type="checkbox"/>	4. referral	4,298	58.63%	2,520	21.68%

Figure 4.10 Sample mediums report showing which channels are sending traffic

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.

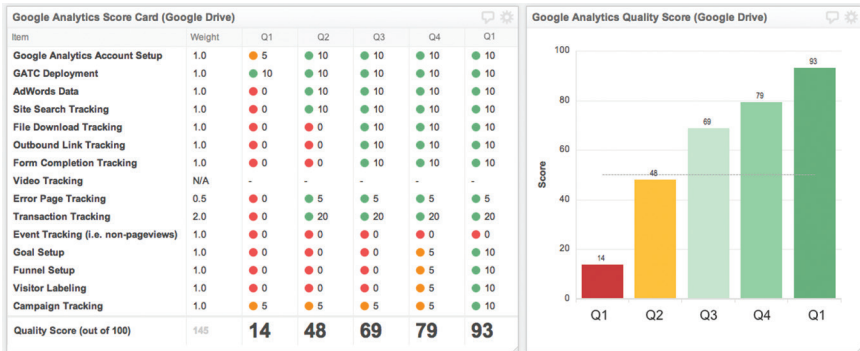


Figure 4.11 Google Analytics quality score tracked over time

SPEAKING OF DATA QUALITY...

When I hear this...

Our stats reports in Facebook.com show much greater numbers than what Google Analytics reports. Which is right?

Who should be assigned to building and maintaining our Google Analytics scorecard?

Who should be assigned to building and maintaining our campaign tracking?

Our quality score varies each month—sometimes as high as 90, sometimes as low as 70. Should we be concerned?

I reply with...

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.

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.

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.

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/entry/cookie_retention_rates.
- 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/>.