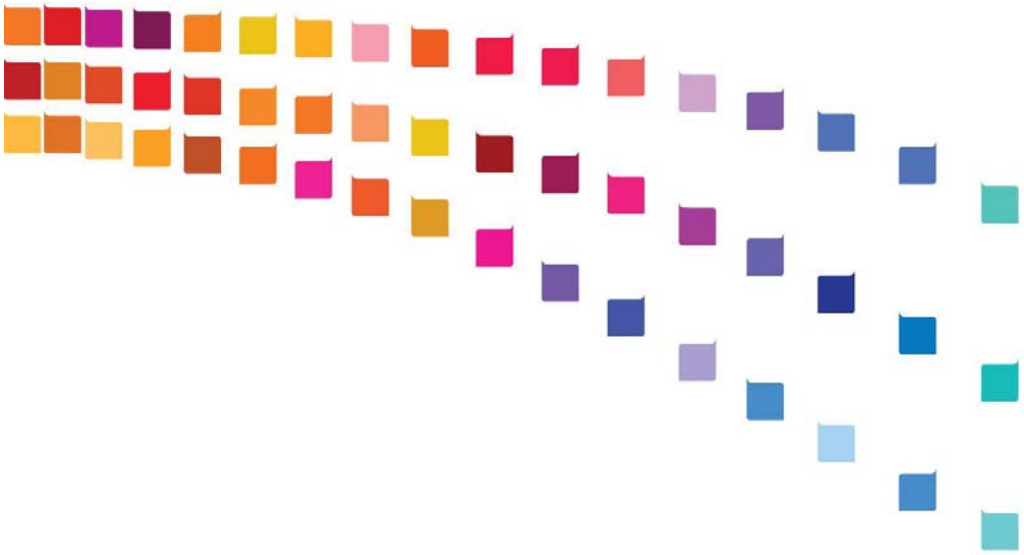
FAQs: Web analytics



FAQs: Web analytics

What is web analytics?

Web analytics is a discipline which developed to permit reporting of activity on websites. Essentially it tells you about how many visitors come to your site, how long they spend and what they do while they are there. It has developed over time to include affiliated analysis activities such as landing page optimisation and layout optimisation via multivariate testing and control cells. The underlying techniques are all taken from traditional business intelligence and data mining processes.

How does web analytics work?

There are a number of methods to obtain visitor data. Originally it was taken from the web-server log, which for server performance reporting kept a record of each page element requested. This approach was superseded because of the introduction of network performance enhancing activities such as page-caching. The result of which was that the true number of page views could no longer been seen at the server. The solution commonly in use today resulted in the process known as page tagging. Page tagging works as follows:

- A page tag is a small piece of Java code. It is executed when the html for the page is loaded in the browser.
- When this Java code executes it adds an activity record into a web-analytics database accessible via the internet and generally hosted by the web analytics provider.

Page tagging although more reliable than web-log analysis creates its own problems, the most significant being the management of the page tags. Ensuring all pages have the appropriate tags requires significant processes and resource. As such new cheaper approaches are coming on stream including single tag solutions and packet sniffing technology.

What data is captured by the page tag?

Typically each page tag data record will consist of the following:

- Date/time, page viewed.
- Some detail on the browser, operating system, hardware platform machine, IP address of the internet connection.
- A browser/machine identification variable known as a cookie ID.
- In addition, the web analytics provider may also provide for the creation and recording of some custom variables, generally these are used for capturing events on the web page, i.e. signing on. These custom variables are often used in the creation of non-standard reports.

The data described above usually results in a large amount of information being recorded. To allow the web-analytics reports to work in real-time, the information in the incoming data records is processed through a number of standardised data cubes. The output from which is of limited value to a traditional CRM analyst. All is not lost however, for those web-analytics providers that charge a licence fee to use their software the source page-tag data is usually available for further processing.

How does web analytics differ from CRM or direct marketing analytics?

Web analytics reports at the aggregate level, i.e. site level. Activities of individual customers, although important, are often rolled up into summary or group level activity. As web analytics has historically been about the aggregate, less attention has been given to the individual. As such, significant work is currently required to convert your typical page-level data into a record of online customer behaviour.

In addition, because of the core design of the http protocol, it is difficult to match every page view with a known individual. A large proportion of a website's user base will not need to identify themselves to use the site. If however these individuals have visited the site in the past on the same machine, using the same browser, then we are able to use the cookie ID as a link variable to join past sessions' activity to the current session. Note, it is becoming more common for cookies to be flushed from a machine more frequently than they used to be cleared. Once the cookies are cleared, the link with the past is lost unless a sign-in process is in place in which case the user ID becomes the common linking variable.

The above implies that we can't map every visitor to a site to a known person and in the cases where we can we can't guarantee that we have all of their online behavioural data accurately assigned to them. The percentages of anonymous users depend on the nature of the site. Retailers will fare better than free media sites for example. Although the anonymous issue is not ideal, the benefits gained from this extra data for known individuals far outweigh the effort involved in collecting it.

Why should CRM practitioners use it?

Reading the FAQ on how online data differs to traditional CRM data, you may question why you should bother trying to use this data at all. Traditionally, CRM analysts have managed to operate effectively with an activity, e.g. a mailing, an outcome and a collection of semi-static and transactional variables to predict likely outcome. Web data could be considered in a similar vein to call centre logs, useful but not essential to customer understanding.

The benefit of web data to customer understanding is in its fine detail. We can see how long a visitor looks at a certain product, how many times they visit the site before buying, what else they do on the site before buying, etc. Using a physical analogy, it is like being able to watch all customers come into a store and record exactly what they do in the store prior to purchase or not. No such detail has ever been available across the whole of a prospect or customer base. The insight lift this data gives you when trying to predict behaviour is significant.

In addition, the proportion of total sales made online has been growing rapidly over the past years. Online is one of the dominant sales and service channels; however the channel, relatively speaking to the others, is still new, best practices are still being developed. Web analytics vendors have not until recently considered customer level data a priority and many CRM analytics practitioners have put getting web analytics data into customer database into the 'too difficult' category. As more sales move online so the need for the CRM practitioner to use this data and interact with this channel will only grow.