Data Mining Queries

Help Documentation

Data Mining Queries

Data Mining - Daily Activity by File

What is the daily activity that this file gets?

This query allows you to select a single page or resource contained within your website and see the daily activity over time. This type of mining is good to identify trends in the activity of that resource, to see if people are accessing the file more or less than they used to.

Terms Used

Bandwidth - Bandwidth represents the total number of kilobytes that were sent to people visiting your site. Bandwidth includes all resources requested by the users.

Hits - A hit represents a request to your web site for a file such as an image, a web page, or a CGI script. One web page may contain several related resources, and as a result, a visitor viewing one web page may trigger several hits. Hits generated as a result of an error (either a 400 or 500 level error) are not counted as actual hits to your site, and are kept separate from successful hits.

Visits - Visits represent the total number of times people have visited your web site. A visit is counted whenever a web site user requests one or more files from the web server. If the user becomes idle for more than a certain amount of time (usually 20 minutes), a new visit is generated when they come back.

Data Mining - Direct Referrals

What are the referrers of this file?

This query is usful for marketing, or to detect image leaching.

For marketing, if you have an image that you have exposed on other web sites that resides on your web server, the URL it was viewed on is recorded every time a browser views your image. This report will list the top URLs or "Refferals" in which your image was viewed.

If you have a nice graphic, chances are that someone else will want to use it on their web site. If they link your image path to their web site, this report will list the top URLs that the image or graphic is viewed from.

Terms Used

Referrer URL - The Referrer URL represents the exact web page from which visitors to your site came

from. A value of "No Referrer" represents a visitor typing your web site directly into the browser, using a bookmark, or following a link from an email client.

Visits - Visits represent the total number of times people have visited your web site. A visit is counted whenever a web site user requests one or more files from the web server. If the user becomes idle for more than a certain amount of time (usually 20 minutes), a new visit is generated when they come back.

Data Mining - Initial Referrals

What sites were people on before they hit this file?

This query reports the original source of visitors that came into your site that eventually hit the page or file you choose. This can be most useful for marketing purposes. For example, if you have a "Thank You" page after a person orders your product, you can identify the original sources of visitors that reached that page, effectively discovering where your orders came from.

Terms Used

Referrer URL - The Referrer URL represents the exact web page from which visitors to your site came from. A value of "No Referrer" represents a visitor typing your web site directly into the browser, using a bookmark, or following a link from an email client.

Visits - Visits represent the total number of times people have visited your web site. A visit is counted whenever a web site user requests one or more files from the web server. If the user becomes idle for more than a certain amount of time (usually 20 minutes), a new visit is generated when they come back.

Data Mining - Visitor Bandwidth

Which visitors used the most bandwidth by downloading this file?

This query allows you to track the bandwidth usage, total bandwidth usage, and the IP address of requests for specific files you have placed on your website. This report is most useful in detecting web site abuse. Typically, you will identify a high-bandwidth file from Top Files or a related report item, then start data mining on it to discover if there are a few IP addresses that are attempting to use your bandwidth by repeatedly downloading files. Those IP addresses can then be metered or denied access on your web server.

Terms Used

Bandwidth - Bandwidth represents the total number of kilobytes that were sent to people visiting your site. Bandwidth includes all resources requested by the users.

Data Mining - Entry Pages

Where did people who hit this file enter the site?

This query allows you to select a single page contained within your website and see the top pages used to access the page. This is a particularly useful form of data mining that allows you to track where people are coming from when they access a specific page. For example, if a page is specified in an advertising campaign or marketing campaign, users can see exactly how people are accessing that page, and verify any information that is given by the particular site that contains the advertising.

Terms Used

Entry Page - The entry page is the first page a web site visitor goes to during a visit

Visits - Visits represent the total number of times people have visited your web site. A visit is counted whenever a web site user requests one or more files from the web server. If the user becomes idle for more than a certain amount of time (usually 20 minutes), a new visit is generated when they come back.

Data Mining - Exit Pages

Where did people exit the site after hitting this file?

This query allows you to select a single page contained within your website and see the last pages people viewed before exiting the website. This is useful in order to track how far users get in a process (such as placing an order), and where they are exiting your site.

Terms Used

Exit Page - The page a visitor leaving your site was on prior to moving to a different website.

Visits - Visits represent the total number of times people have visited your web site. A visit is counted whenever a web site user requests one or more files from the web server. If the user becomes idle for more than a certain amount of time (usually 20 minutes), a new visit is generated when they come back.

Data Mining - Repeat Visits

How many times did people return to my site before hitting this file?

This query allows you to select a single page contained within your website and see the number of times a person visited the site prior to reaching and viewing the page selected. The report itself shows

the IP address of the visitor, the initial Entry Page for that visitor before they viewed the page selected, and the number of visits before the page selected was viewed.

Terms Used

Entry Page - The entry page is the first page a web site visitor goes to during a visit

Visits - Visits represent the total number of times people have visited your web site. A visit is counted whenever a web site user requests one or more files from the web server. If the user becomes idle for more than a certain amount of time (usually 20 minutes), a new visit is generated when they come back.

Data Mining - Query Strings

What query strings were passed to this file?

This query allows you to view the various query strings that are passed to a particular page of your website.

A typical example is the SmarterStats Help system. Whenever anyone clicks on the "Help" graphic or a "What is this?" link within a specific Report Item in the Web Interface, a query string is passed from the SmarterStats application to the SmarterTools online help. This query string relates to the particular page or Report Item the user is seeking help with. This report shows what query string was passed to generate the page.

This query differs from the Query Items query in that it does not separate query string items. For example, assuming you have a hit that looks like /help/default.aspx?section=Administrator&page=Getting+Started . The Query Items report will show two lines: One for section=Administrator and one for page=Getting+Started . Conversely, the Query Strings report will show one line that looks like section=Administrator&page=Getting+Started .

Terms Used

Hits - A hit represents a request to your web site for a file such as an image, a web page, or a CGI script. One web page may contain several related resources, and as a result, a visitor viewing one web page may trigger several hits. Hits generated as a result of an error (either a 400 or 500 level error) are not counted as actual hits to your site, and are kept separate from successful hits.

Query String - Query strings are parameters that are passed to web pages in the URL string, commonly scripted pages.

Data Mining - Query Items

What query items were passed to this file?

This query allows you to view the various query string items that are passed to a particular page of your website.

A typical example is the SmarterStats Help system. Whenever anyone clicks on the "Help" graphic or a "What is this?" link within a specific Report Item, a query string is passed from the SmarterStats application to the SmarterTools online help. This query string relates to the particular page or Report Item the user is seeking help with. This report shows what individual items were passed to generate the page.

This query differs from the Query Strings query in that it separates each query string item and reports them individually. For example, assuming you have a hit that looks like /help/default.aspx?section=Administrator&page=Getting+Started . The Query Items report will show two lines: One for section=Administrator and one for page=Getting+Started . Conversely, the Query Strings report will show one line that looks like section=Administrator&page=Getting+Started .

Terms Used

Hits - A hit represents a request to your web site for a file such as an image, a web page, or a CGI script. One web page may contain several related resources, and as a result, a visitor viewing one web page may trigger several hits. Hits generated as a result of an error (either a 400 or 500 level error) are not counted as actual hits to your site, and are kept separate from successful hits.

Query Item - A query item is an individual parameter passed to a web page, typically a script.