Topic 2 : Clickstream Analysis I

ST2228 WEB AND MOBILE ANALYTICS

Learning Objectives

- Clickstream Analysis
  - Distinguish between metrics and dimensions
  - Formulate the measurement plan and web analytics strategy
  - Plan suitable micro and macro conversions
  - Measure the various KPIs used
  - Analyse how users move from one page to another

What is Clickstream Analytics?

- Recall:
  - A clickstream is the recording of the parts of the screen a computer user clicks on while web browsing.
  - As the user clicks anywhere in the webpage or website, the action is logged on a web browser or inside the web server.
  - A clickstream is a series of page requests and every page requested generates a signal.

Basic Web tech terms

WHAT IS URL?

WHAT IS URL PARAMETER?

WHAT IS COOKIE?

URL

1. Preparing to understand basics:
   - URLs
     - *Why should you care:*
       - Important as it identifies a set of data that is connected to that specific domain name
     - *What you should care about:*
       - **What is the identity of your website?** (www.sp.edu.sg, www.sb.sp.edu.sg) This will reveal your data strategy, customer session, cookie strategy and ability to get views for your website data.
       - **What can the URL tell about your site structure?** Intelligence of your URL structure will determine how much work needs to be done by analytics team.

URL Parameter

- URL Parameter
  - *What is it?*
    - http://www.google.com/support/forum/p/apps-apis/thread?tid=440427120aed9a2b&hl=en
    - Anything after the "?" is called a URL parameter
  - *Why should you care:*
    - Used to drive dynamic experiences on web pages
    - Used for tracking purposes

Cookie

- Cookies
  - Session cookies, persistent cookies
  - *Why should you care?*
    - Important in creating personalized experiences for website visitors
    - Useful for tracking repeat visits
    - Useful for storing and reporting interesting and useful information

Cookie

- Cookies
  - *What should you care about?*

  - Document exactly what cookies are being set
  - Your most important cookies should be first party cookies so that they stand a chance of being accepted by high-security settings. 3rd party cookies are usually blown away by anti-spyware software
  - Measure the cookie rejection and deletion rates of your website visitors ➔ help to measure how good/bad web analytics data

Data Collection

CLICKSTREAM

Data Collection

Data Collection

- GIGO = Garbage In Garbage Out
- Computers have gotten *slightly* smarter to make sense of garbage **BUT** they can only do so much ➔ Input data is still very important
- Data capture mechanisms are still in a state of flux as they try to keep pace with the changing nature of the web

Data Collection

- Understanding the Data Landscape
  - There are a number of ways to collect data as a customer interacts with our website (e.g. web log files, JavaScript tags, packet sniffers etc.)

Be **thorough** as some mechanisms are like this(e.g. miss a Javascript tag on page and miss out important data)

Data Collection

- Understanding the Data Landscape
  - Many industry-leading tools require options to be thought through up front and explicit decisions to be made on what data to capture
    - If the correct data is not captured at the beginning, you won't be able to do the analysis
  - Sometimes you may need more than one method of data collection
    - JavaScript tags to collect website behavior and web logs to analyze the behavior of robots on the website (because search engine robots do not execute JavaScript tags)

Data Collection

- Understanding the Data Landscape
  - Competitive data is important!
    - Need to understand not just how your own company is doing BUT also how your competitors are faring
  - Privacy issues
    - As mechanisms are put in place to capture data, it is VERY IMPORTANT that you declare what data is being captured
    - Customers are extremely concerned about their privacy

Data Collection

Clickstream Data

- Clickstream:
  - Virtual trail that user leaves behind while surfing
  - Record of the user's activity on the Internet (every website and every page, how long user was on the page, what order etc.)
  - Can be captured by ISPs or individual websites

Data Collection

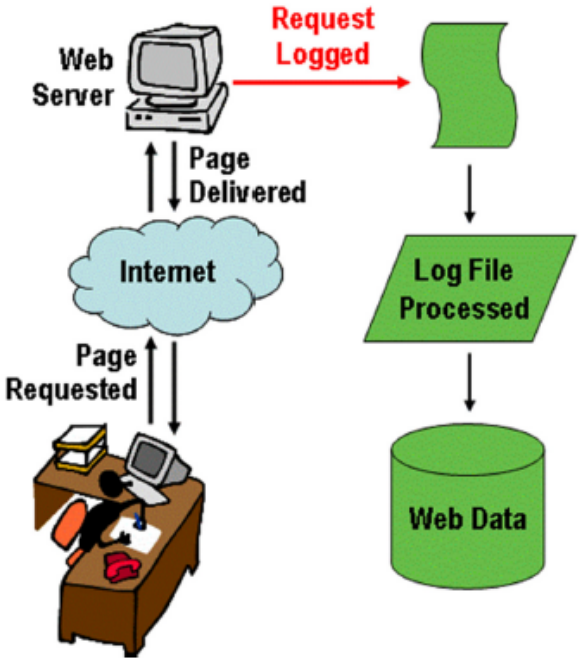
4 main ways to capture clickstream data:

- Web logs
- Web beacons
- JavaScript tags
- Packet sniffing

Data Collection

Clickstream Data

- Web Logs
  - Original source of data collection
  - Originally developed to capture errors generated by web servers
  - Has evolved to capture more data to handle marketing questions



Data Collection

Clickstream Data

- Web Logs
  - Benefits:
    - Most easily accessible source of data
    - Many log parsers available for free
    - Only capture mechanism that will capture and store the visits and behavior of search engine robots on website (Search engine robots don't execute JavaScript tags ➔ if you want to analyze visits by Google/Yahoo search engine robots to ensure website is being crawled and indexed correctly, HAVE to use web logs)
    - Own the data

Data Collection

Clickstream Data

- Web Logs
  - Concerns:
    - Primarily geared towards technical information, not so much for business/marketing information
    - May need to depend on the technical/IT team to make changes
    - Web logs created to capture all hits on the server ➔ have to be careful to apply the right filters to remove unnecessary info
    - Page caching can mean that some of the traffic is invisible to you because websites can be cached at the ISP or proxy servers ➔ when someone from the ISP's network requests for homepage, it is served from ISP and NOT your web server

Data Collection

Clickstream Data

- Web Logs
  - Recommendation:
    - Use web logs to analyze search engine robots behavior to measure success of search engine optimization efforts
    - Use web logs to complement data captured using other methodologies

Data Collection

Clickstream Data

- Web Beacons
  - Developed as a prime way to capture eyeballs
  - Usually 1 x 1 pixel transparent images placed in web pages within an img src HTML tag and hosted on a 3rd party server
  - 3rd party server will send the image back to the browser along with code that can read cookies and capture anonymous visitor data such as:
    - The fact the page was viewed
    - The IP address
    - Time the page was viewed
    - Cookies that were previously set, etc.

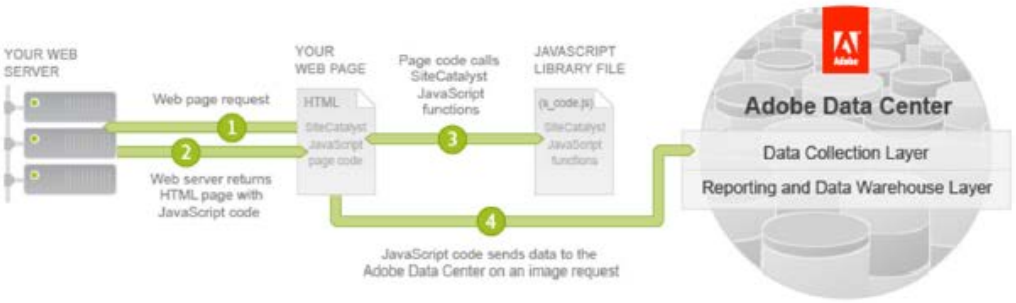
Data Collection

As each web page is requested in a browser, your web server returns the HTML page with the embedded JavaScript page code (see Figure 2). The JavaScript page code sets the values for analytic data that you are collecting and calls SiteCatalyst-specific functions and global variables in the JavaScript library file.

Figure 2. An overview of the SiteCatalyst JavaScript Tagging code

The SiteCatalyst JavaScript Tagging code builds an image request for a 1x1 pixel image, also called a web beacon, that concatenates a query string of name/value pairs of analytics data to send to the Adobe Data Collection Layer (see Figure 3).

The web beacon image, which is just a dummy request to pass the data, is returned to your web page, but is not displayed.



Data Collection

Clickstream Data

- Web Beacons
  - Also used in emails
  - When email loads in email reader, data is sent back and recorded
  - However with JavaScript becoming more prevalent, web beacons are not as popular and are mostly used instead to track basics around banner ads and emails

Data Collection

Clickstream Data

- Web Beacons
  - Benefits
    - Easy to implement because they are just a couple of lines of codes
    - Can optimize exactly what data the beacon collects ➔ keep logs manageable and won't require complex filtering
    - Good for collecting data across multiple websites or domains
      - E.g. if you are a publisher with content on many sites, beacons can be used to collect and store data from all these sites on one server (the one sending out all the data collection requests)

Data Collection

Clickstream Data

- Web Beacons
  - Concerns
    - Have a bad reputation because usually associated with advertising ➔ antispyware programs automatically remove cookies which hampers ability to collect the data
    - Beacons are not as customizable as JavaScript tags in terms of the data they can capture
      - Capture less data but can do so across a broad range of websites
    - If image requests are turned off (eg. In emails), data can't be collected

Data Collection

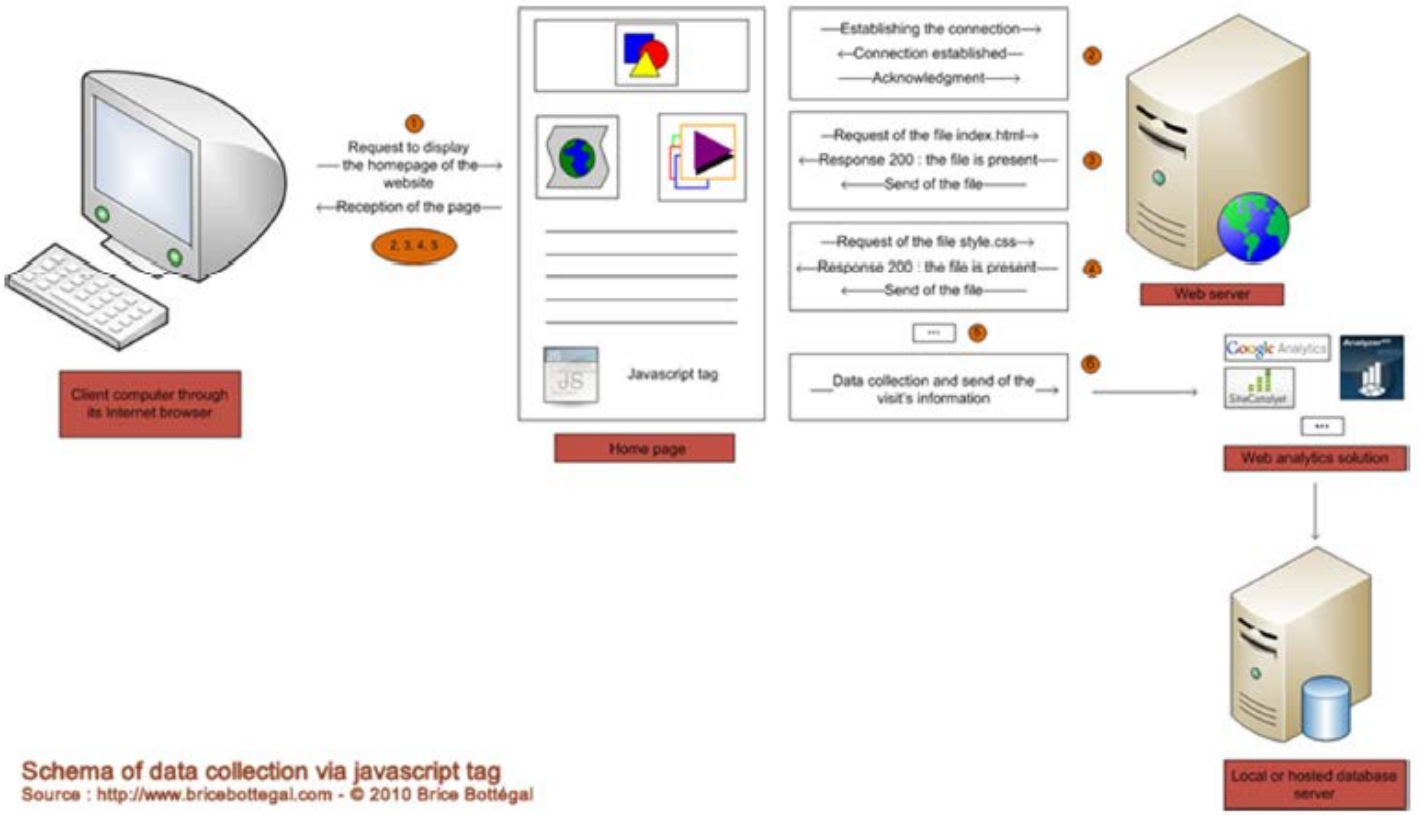
Clickstream Data

- Web Beacons
  - Recommendation
    - Best used when want to track visitor behavior across multiple websites or track email open/view rates
    - For rich website analysis, may still have to rely on other methods because data captured is not as rich

Data Collection

- JavaScript tags
  - Allows for more data to be captured more accurately
  - Data serving was separated from data capture
    - Reduces the reliance on corporate IT departments for various data capture requests
    - Data capture can be moved to 3rd party analytics vendors
    - Companies did not need to host their own infrastructure to collect data

Data Collection



Schema of data collection via javascript tag
Source : http://www.bricebottegal.com - © 2010 Brice Bottégal

Data Collection

Clickstream Data

- JavaScript tags
  - Benefits
    - One of the easiest initial implementations (next to web logs)
    - Best choice if you don't have access to the web server
      - Can install tags easily in your pages and use a vendor to do your reporting (Best for SMEs)
    - Page caching is not a problem because regardless of where the page is served from, the tag will execute
    - Have great deal of control over exactly what data is collected
    - Able to separate data capture from data serving
    - Tracking across multiple domains becomes easier because 3rd party cookie and its identifying elements stay consistent

Data Collection

Clickstream Data

- JavaScript tags
  - Concerns
    - Not all website visitors have JavaScript turned on (because of privacy)
    - Data collected from JavaScript tagging is divorced from other metadata
    - JavaScript tags collect data "browser side" and NOT "server side"
      - Some websites store data on servers during visitor sessions ➔ tags will not capture that data
    - Capturing data in JavaScript tags about downloads and redirects is harder than with web logs
    - If website is JavaScript heavy, there may be conflicts

Data Collection

Clickstream Data

- JavaScript tags
  - Recommendation
    - Best when you need to control the data being collected
    - Just that you will need to leverage on web logs to measure search engine optimization (SEO) on your website

Data Collection

Clickstream Data

- Packet Sniffing
  - One of the most sophisticated ways
  - Is a piece of software or hardware capable of monitoring all network traffic.
  - Able to capture all incoming and outgoing traffic (e.g. clear-text passwords, user names and other private or sensitive details)

Data Collection

Clickstream Data

- Packet Sniffing
  - Benefits
    - Eliminates the need for JavaScript tags
    - Faster to implement than most other methods except JavaScript tagging
    - Can collect lots of data instantly
      - Much more than JavaScript tagging
      - Most comprehensive way of collecting data
    - Have the flexibility of using first party for cookies

Data Collection

Clickstream Data

- Packet Sniffing
  - Concerns
    - May encounter resistance from IT dept because it involves adding another non-standard software on web servers / physical server in their highly-secure data center
    - Will need to filter out just the needed data from the large amounts of raw data
    - Can result in privacy issues
    - Packet sniffer does not get any data for cached pages because no request comes to the web server (so will need to use JavaScript tags to complement)
    - Can be expensive if you have many web servers sitting on many networks because will need to install on ALL the networks

Topic 2 : Clickstream Analysis II

ST2228 WEB AND MOBILE ANALYTICS

Learning Objectives

- Exploring Google Analytics
  - Interface Navigation
  - Tabs
    - Visitors
    - Traffic Sources
    - Content
    - Goals
  - Analysing Contextual Comparison and Trends
  - Customizing Tracking and Reports
  - Predefined and Custom Filters

Getting Started with GA

- Checklist
  - Google Account
  - Blogger Account
  - Access to Google Analytics
  - Create a Blog on Blogger
  - Add Google Analytics to Blog

http://www.youtube.com/watch?v=mLWuRyVNte8

Key Metrics

Metric	Description
% exit	% of users who exit from a page as a share of total pageviews.
Bounce rate	% of visits in which the visitor enters and exits on the same page, without visiting any other pages on the site.
Click	Single instance of a visitor clicking on a link from one page to another on the same site.
Click path	Sequence of clicks that visitors follow on a given path.
Click-through rate (CTR)	% of visitors who view an item and then click it. Calculated by dividing the number of clicks by the number of impressions.

Key Metrics

Metric	Description
Depth of visit (pageviews per session)	Average number of pageviews a visitor initiates before ending his session. Calculated by dividing the total number of pageviews by the total number of sessions.
First visit	First visit from visitor who has not made previous visits to the site.
Hit	Request for a file from web server. A hit is not the same as pageview. Example, a web page that has four images will result in five hits to the server.
Impression	Single display of an advertisement on a web page.
Loyalty	Measurement of how often visitors come to a site. Calculated by dividing the total number of sessions or visits by the total number of unique visitors.

Key Metrics

Metric	Description
Pageview	Display of a complete web page. One visitor looking at a single page of your site generates one pageview.
Session	Series of pageviews from the same visitor with no more than 30 minutes between pageviews and with no visits to other sites between pageviews. Unlike visits, a session ends when a visitor opens a page on another site.
Unique visitor	Visitor who visits your site one or more times within a given timeframe, typically a single 24-hr period. A visitor can make multiple visits during the timeframe but this counts as just a single unique visitor.
Visit	Series of pageviews from the same visitor with no more than 30 minutes between each pageview. Unlike a session, a visit continues (for 30 minutes) even after a visitor leaves your site.

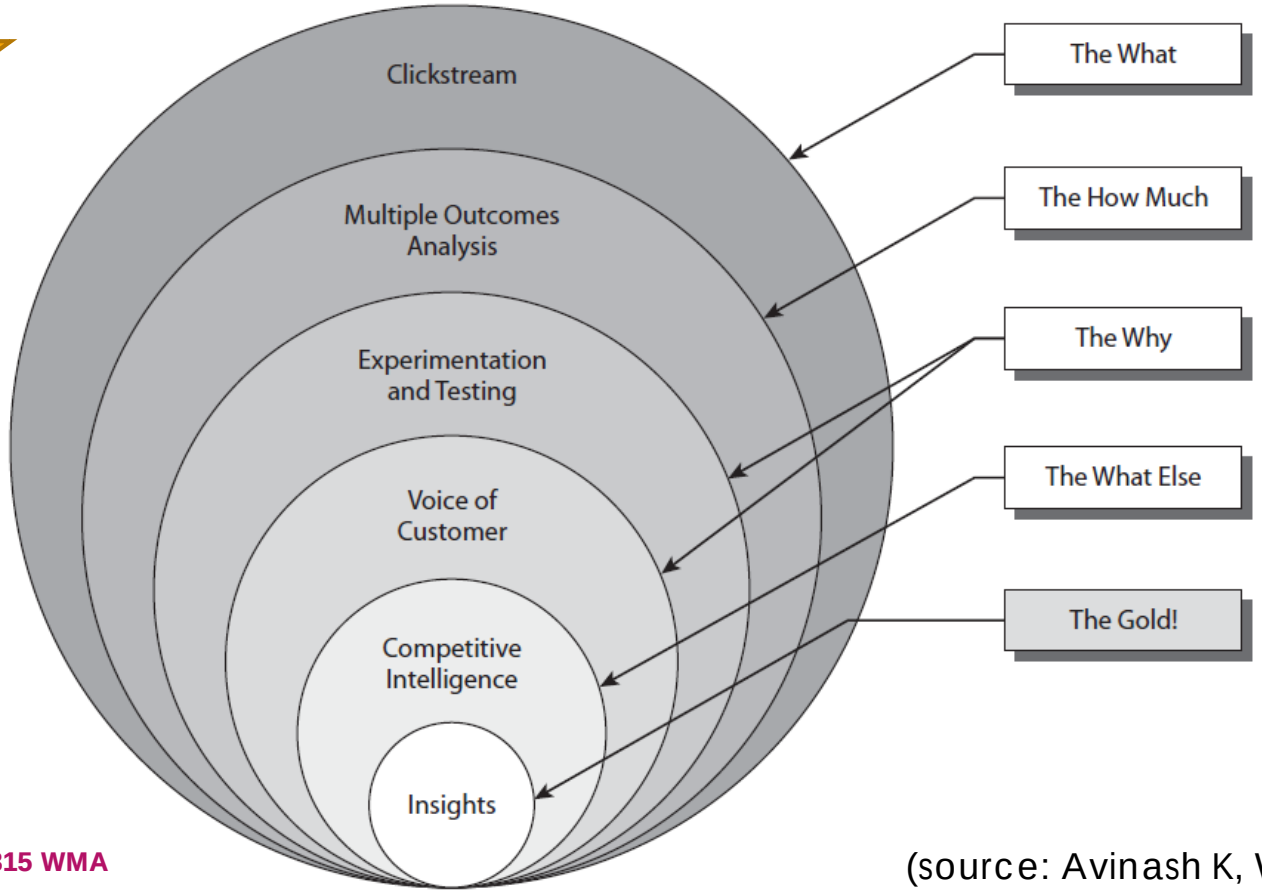
Interjection

KEEP YOUR HYPOTHESIS IN MIND

What is the Question?

Mouse Loves Rice?

Key questions for WA

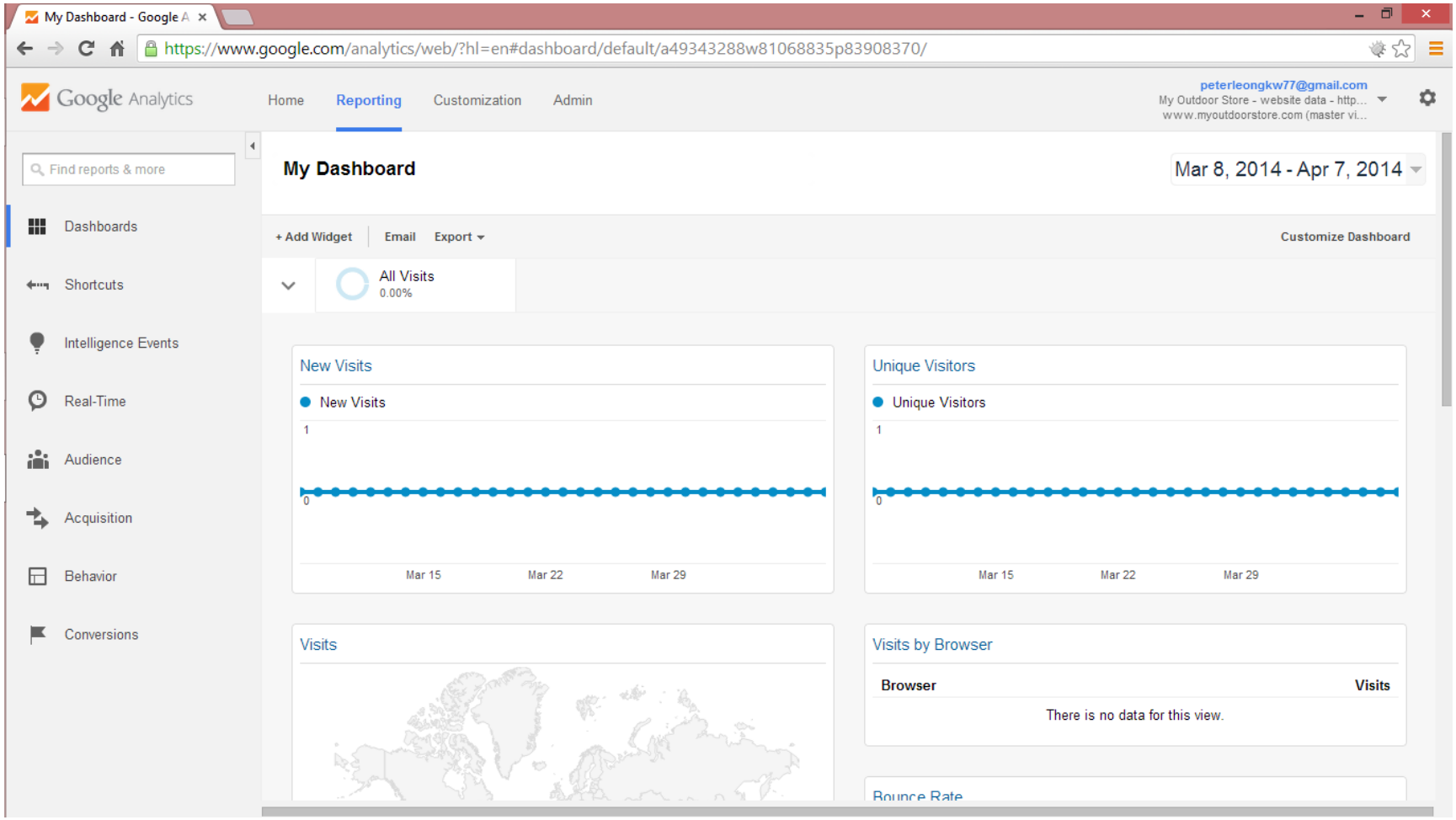


(source: Avinash K, Web Analytics 2.0)

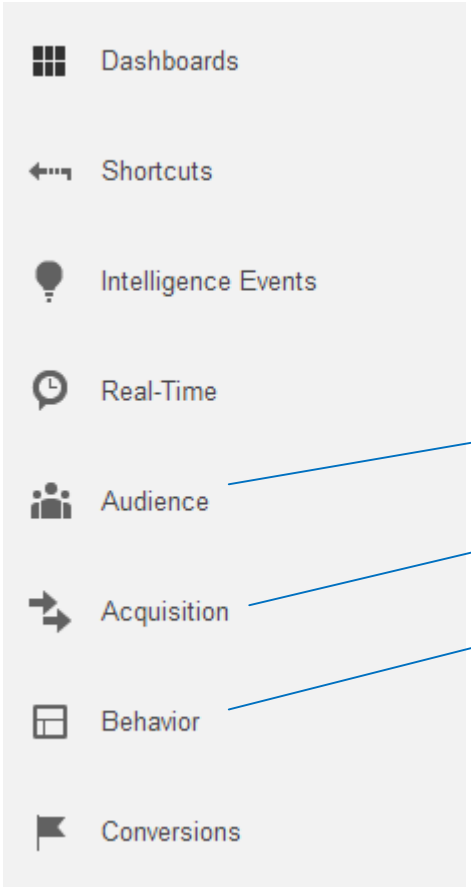
Getting to know GA

1) Unusually comprehensive in the metrics it tracks
2) Relatively easy to use
3) Completely free
4) Popular with websites both large and small
5) Uses onsite analytics to track visitor behaviour
6) Generates a unique JavaScript code for site to be tracked

GA Product Tour



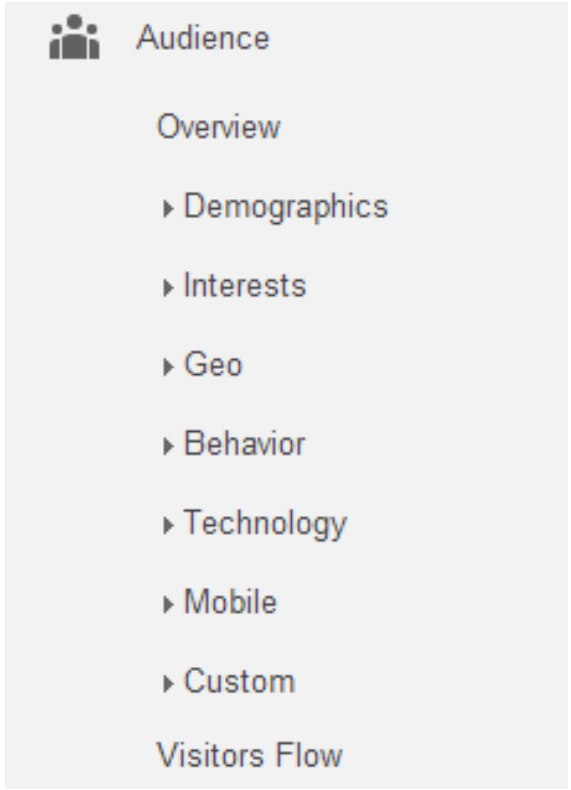
Key Features



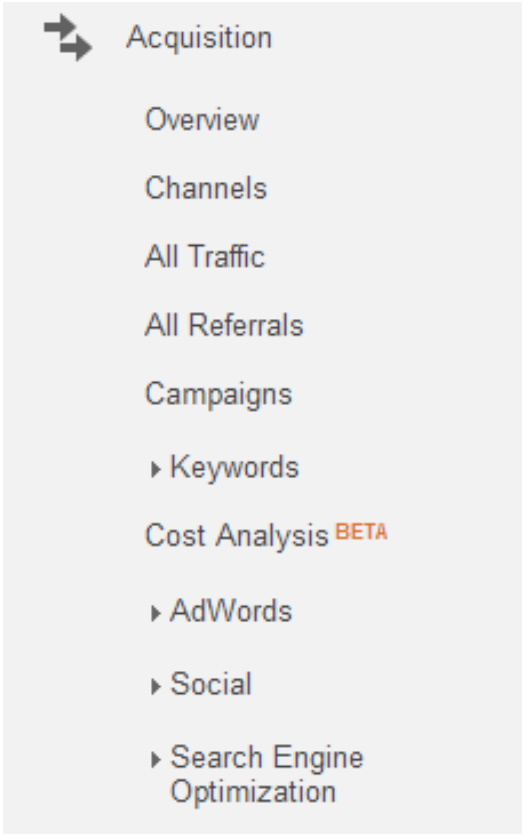
Old GA main menu

1) Dashboard
2) Intelligence
3) Visitors
4) Traffic Sources
5) Content
6) Goals
7) E-Commerce

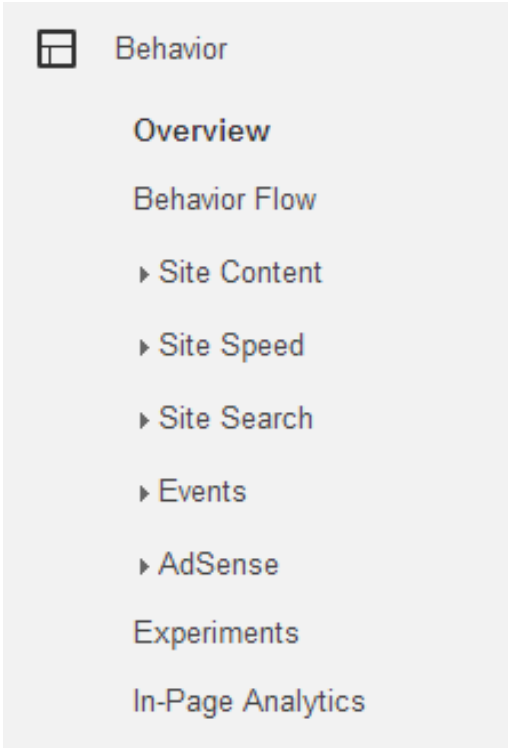
Key Features: Audience



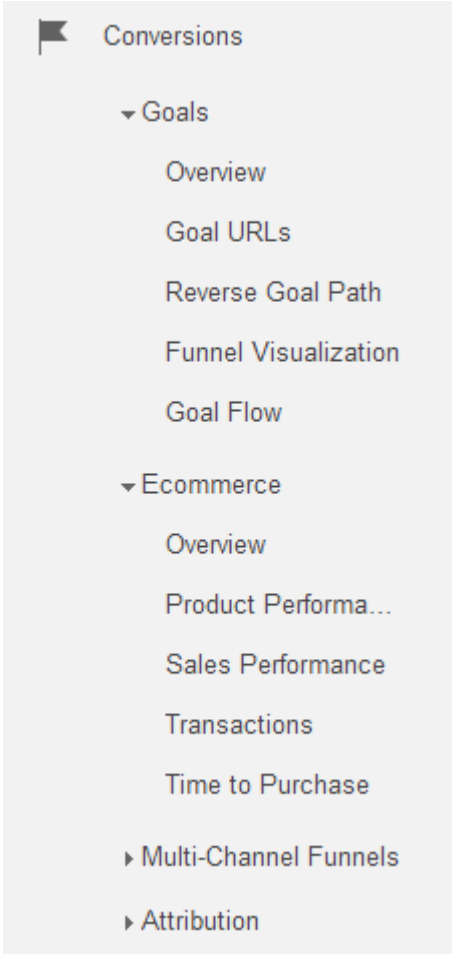
Key Features: Acquisition



Key Features: Behavior



Key Features: Conversion



Interface Navigation

- http://www.youtube.com/watch?v=xyh8iG5mRls
- Exercise

- Log into your Google Analytics account
- Export the dashboard data of your blog into a pdf file and save it

Audience

- http://www.youtube.com/watch?v=ORrYEEH_KPc

Exercise – Map Overlay of % New Visits

- Go to the tab, Visitors -> Map Overlay
- Change the matrix of the map to % New Visits
- Export the Map Overlay data into a pdf and save it

Acquisition

- http://www.youtube.com/watch?v=ii0T5JUO2BY
- Exercise

Exercise – Comparing All Traffic Sources

- Go to the tab, Traffic Sources -> All Traffic Sources
- Click on the Comparison Icon
- Export the All Traffic Sources data into a pdf and save it

Behavior

- http://www.youtube.com/watch?v=Dz-zgq6OqTI

Exercise – Navigation Summary

- Go to the tab, Content -> Overview
- Click on the Navigation Summary link
- Export the Navigation Summary data into a pdf and save it

Custom Report & Dashboards

- http://www.youtube.com/watch?v=l1b4GT-GuEs

Activity

Exercise – Setting a Predefined Segment

- Go to the tab, Content -> Overview
- Following the steps shown in the video, select New Visitors and Direct Traffic as your predefined segments
- Export the Content Overview data into a pdf and save it

Activity

Exercise – Setting a Custom Segment

- Go to the tab, Content -> Overview
- Following the steps shown in the video, create a new Advanced (Custom) Segment with the following values:
  - Dimensions
    - ✓ Mobile – Yes (Matches exactly)
  - Metrics
    - ✓ Visits – 1 (Greater than)
- Name your segment as "Visits Using Mobile" and save it
- Apply the newly created segment to the Content Overview report
- Export the Content Overview data into a pdf and save it

Filters

- http://www.youtube.com/watch?v=dzwRzUEc_tA

Exercise – Applying Filter to Exclude Internal Traffic

- Following the steps shown in the video, use the predefined filter to Exclude Internal Traffic from your Website Profile
- Go to http://whatismyipaddress.com to find out your IP address

Conversion

Conversion rate = $\frac{\text{Outcomes}}{\text{Number of Unique Visitors}}$ for a **given time period**.

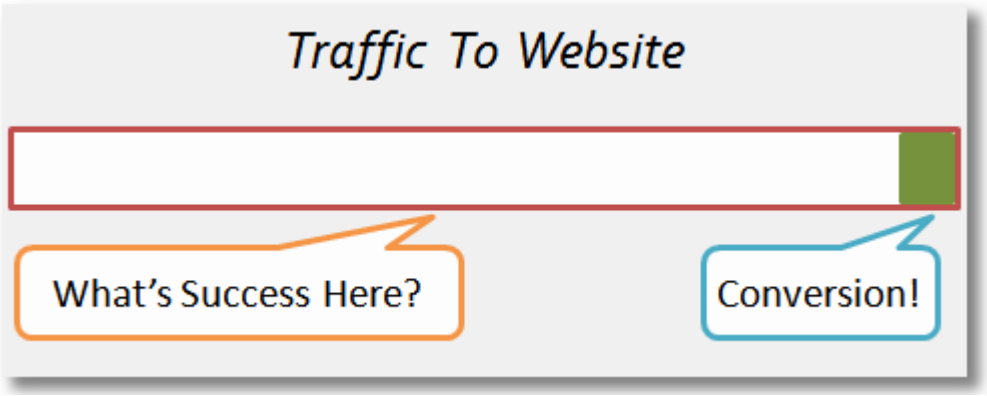
- Outcomes – Any reason why website exist:
  - Total number of orders submitted or total number of leads collected or total number of newsletter/email sign-ups.
  - Number of people who completed a task, so for a support site the number of people who got to a FAQ or a answer or a knowledge base article.

Adapted from http://www.kaushik.net/avinash/excellent-analytics-tip5-conversion-rate-basics-best-practices/

Conversion

- *Time period*:
  - If you are measuring Weekly Conversion Rate it is the sum of orders during that time and sum of unique shopper_id's during the same weekly time period.
  - It is **not** the sum of unique visitors on each day and then a sum of that daily unique visitor number for a week.
  - For Monthly sum of orders during that month and sum of unique shopper_id's during that time.
  - It is **not** recommended to sum daily unique visitors to get a total for month or week.

Conversion

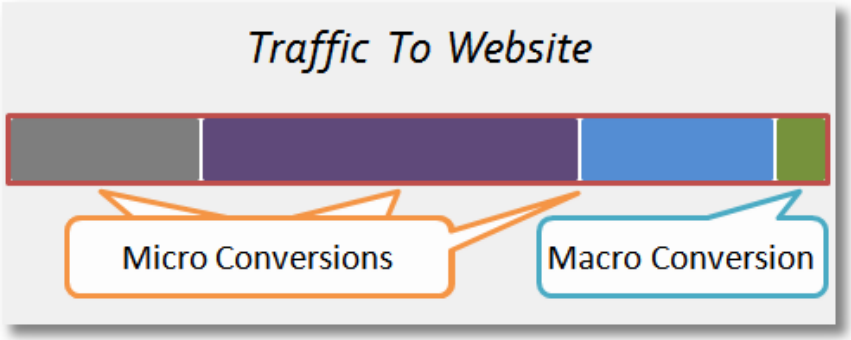


- People don't just come to a site to Buy.
- They are there to Research products and services (and buy offline).
- They are looking to get Support.
- They might be there to look at your latest Blog Post, etc.

Adapted from http://www.kaushik.net/avinash/excellent-analytics-tip-13-measure-macro-and-micro-conversions/

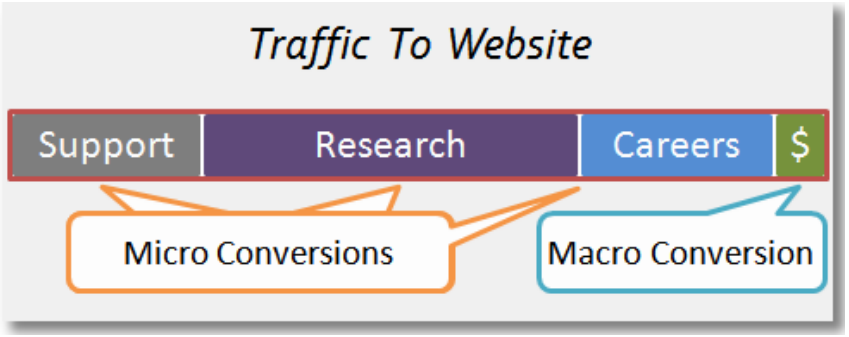
Conversion

From this



- Micro-conversions: each of these tasks.

To this



Adapted from http://www.kaushik.net/avinash/excellent-analytics-tip-13-measure-macro-and-micro-conversions/

Conversion

Benefits of measuring micro conversions:

- You'll focus on more than just the main reason the site was created.
- You'll measure multi-channel impact, well beyond your website.
- It will force you to understand the multiple persona's on your website.
- It will encourage you to segment visitors and visits and behavior and outcomes.
- You'll realize the limits of a pure clickstream strategy.
- By expanding your measurement horizon and seeking insights from a broader area means you'll know what to do with all this data.