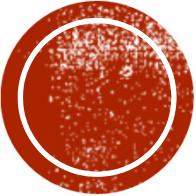


WEB ANALYTICS

PROF. XINXIN LI





I. INTRO TO WEB ANALYTICS



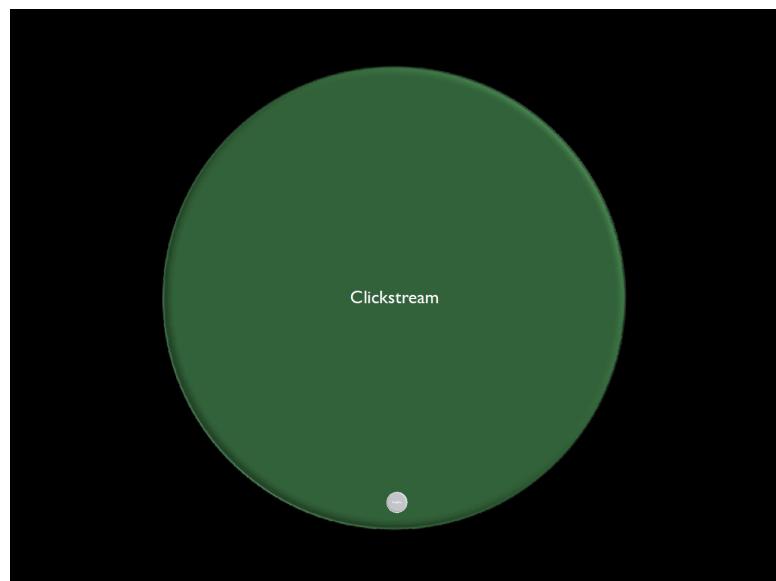
MOTIVATION

- The web is a massive collection of:
 - Documents, data, images, videos
 - Hyper-links
 - Transaction data
 - Behavioral data, etc.
- Discovering interesting patterns leads to
 - better information and knowledge acquisition
 - business intelligence
 - more efficient and effective Web strategies

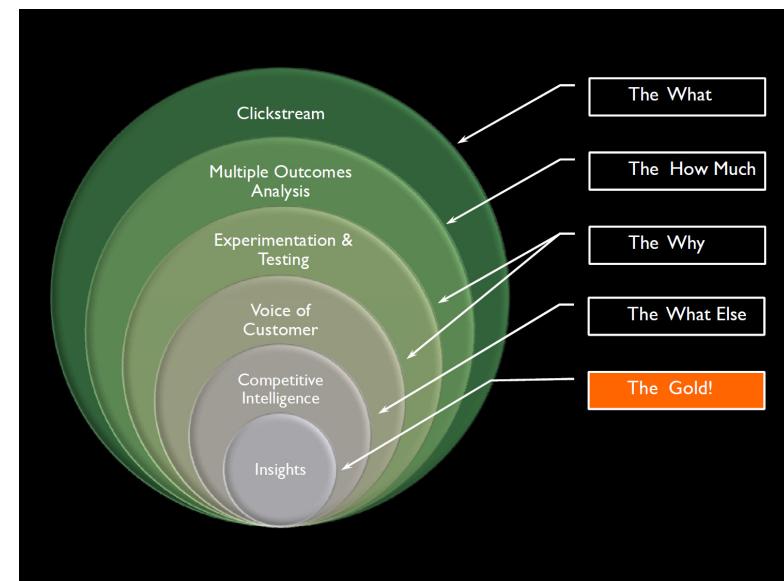
WHAT IS WEB ANALYTICS?

- **Web analytics** is the **collection, analysis and reporting** of internet data for purposes of understanding and **optimizing** web usage which translates into **desired outcome** (from Wikipedia)

WEB ANALYTICS EVOLUTION



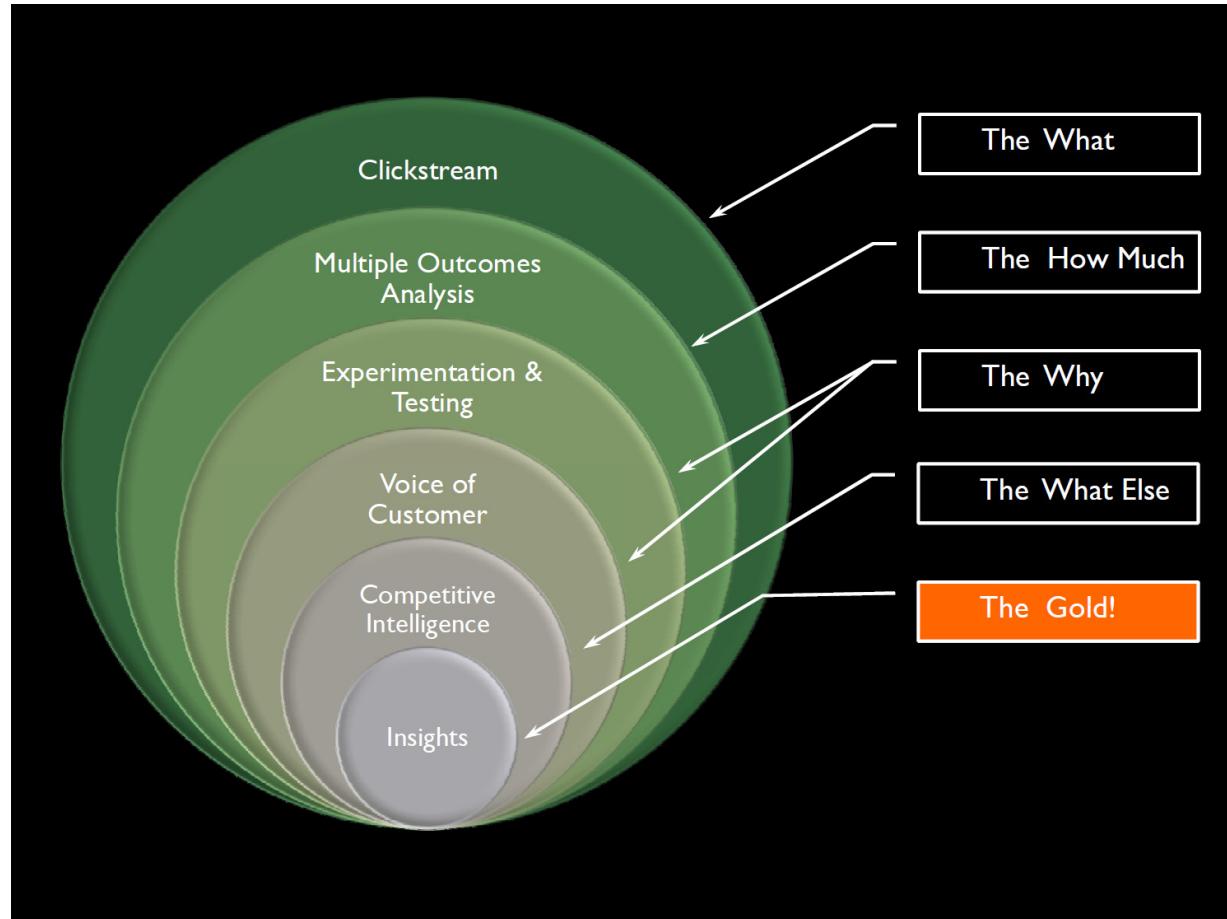
Web Analytics 1.0



Web Analytics 2.0

<https://www.kaushik.net/avinash>

WEB ANALYTICS 2.0



The What

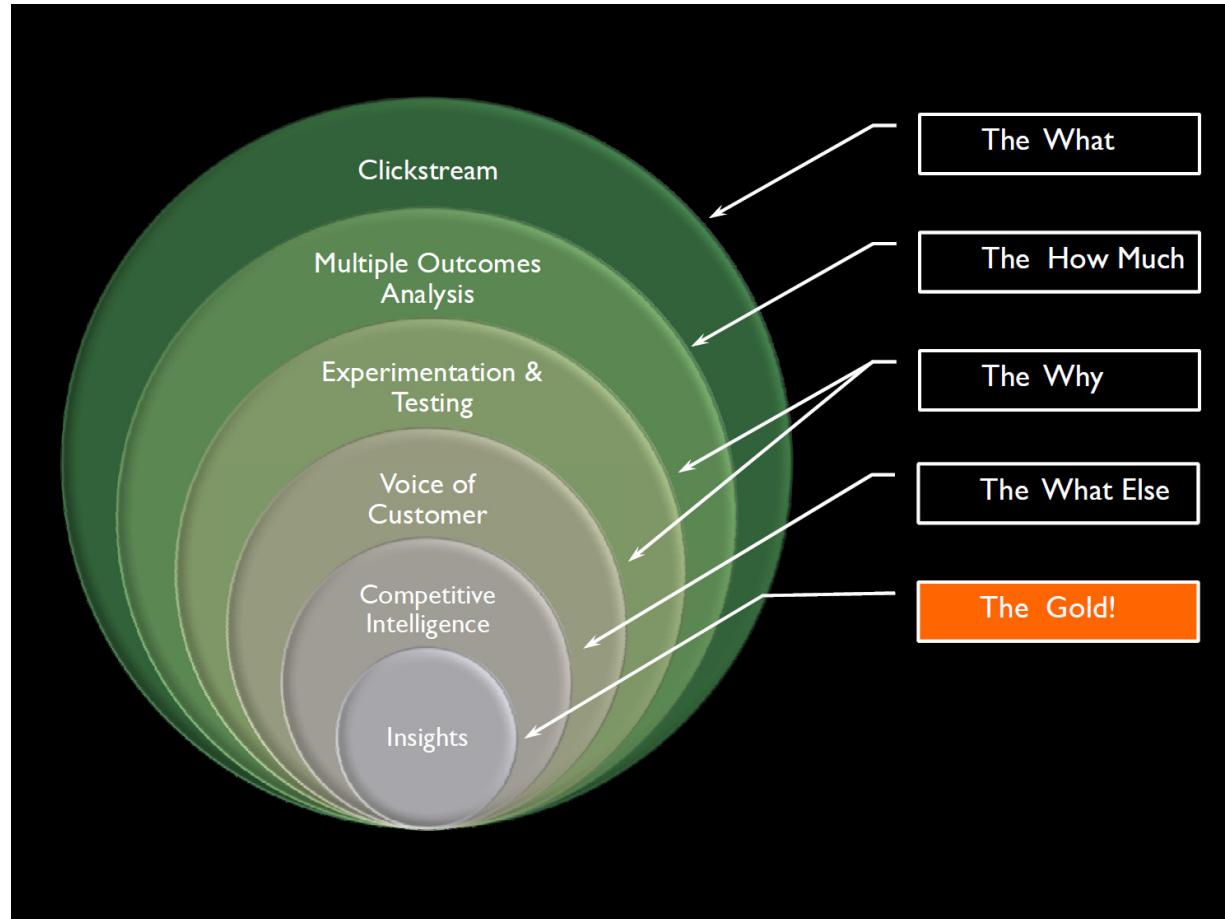
- Clickstream: collect, store, process and analyze web usage data
- Foundational data to analyze pages and campaigns
 - What pages did people view?
 - What products did people purchase?
 - What was the average time spent?
 - What sources did they come from?
 - etc.

AN EXAMPLE OF CLICKSTREAM DATA

A “stream” of clicks or sequences of user browsing behavior

machine_id	event_time	page_viewed
16400000	25-Apr-09 09:52:01	https://msbapm.business.uconn.edu/
16400000	25-Apr-09 10:07:30	https://msbapm.business.uconn.edu/about/faq/
16400000	25-Apr-09 10:25:56	https://msbapm.business.uconn.edu/about/faq/#application
16400000	25-Apr-09 10:31:55	https://www.business.uconn.edu/contact/ms-in-business-analytics-and-project-management/
17900000	25-Apr-09 14:40:58	https://msbapm.business.uconn.edu/admissions/requirements/
17900000	25-Apr-09 14:59:35	https://grad.business.uconn.edu/events/hartford/
17900000	25-Apr-09 15:10:34	https://connect.grad.uconn.edu/register/?id=cd89fa62-7618-44f4-9bb1-1c2a454b7730

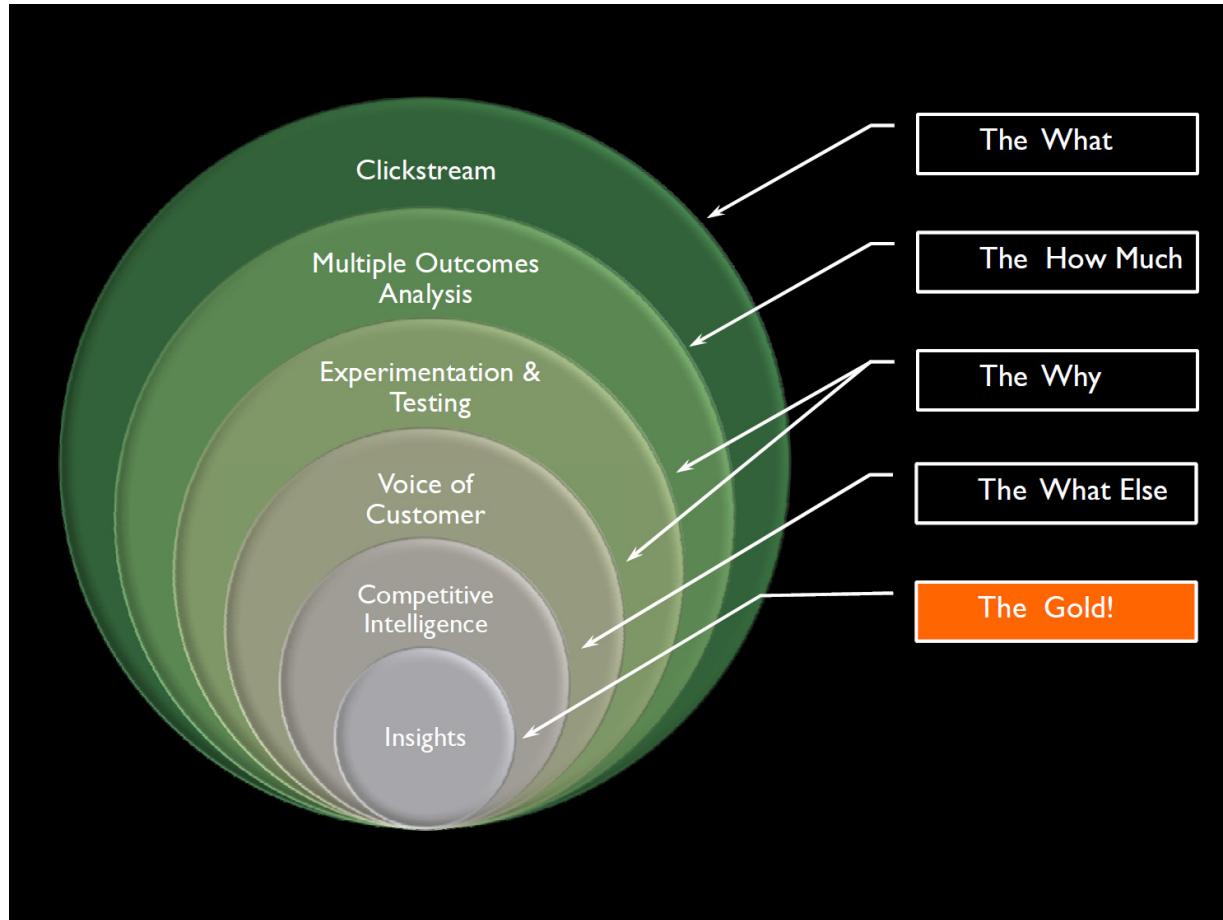
WEB ANALYTICS 2.0



The How much?

- Connect the customer behavior to the bottom line
- Analyze by how much outcomes change:
 - Has revenue been increased?
 - Has cost been reduced?
 - Has customer satisfaction/loyalty gone up?

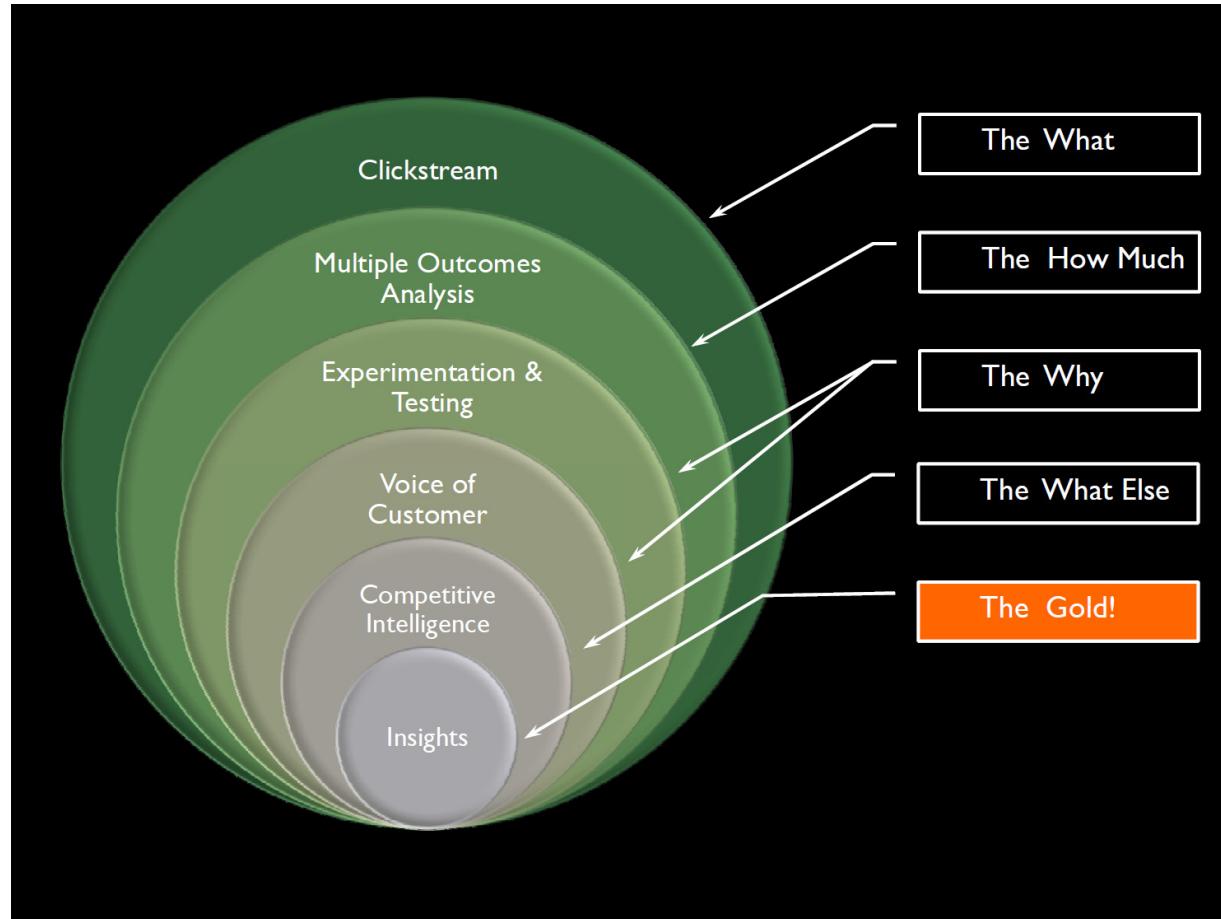
WEB ANALYTICS 2.0



The Why

- Leveraging the power of experimentation and testing tools
- Voice of customers from surveys, e.g.,
 - why do people abandon a website?
 - why are visitors not buying any product?

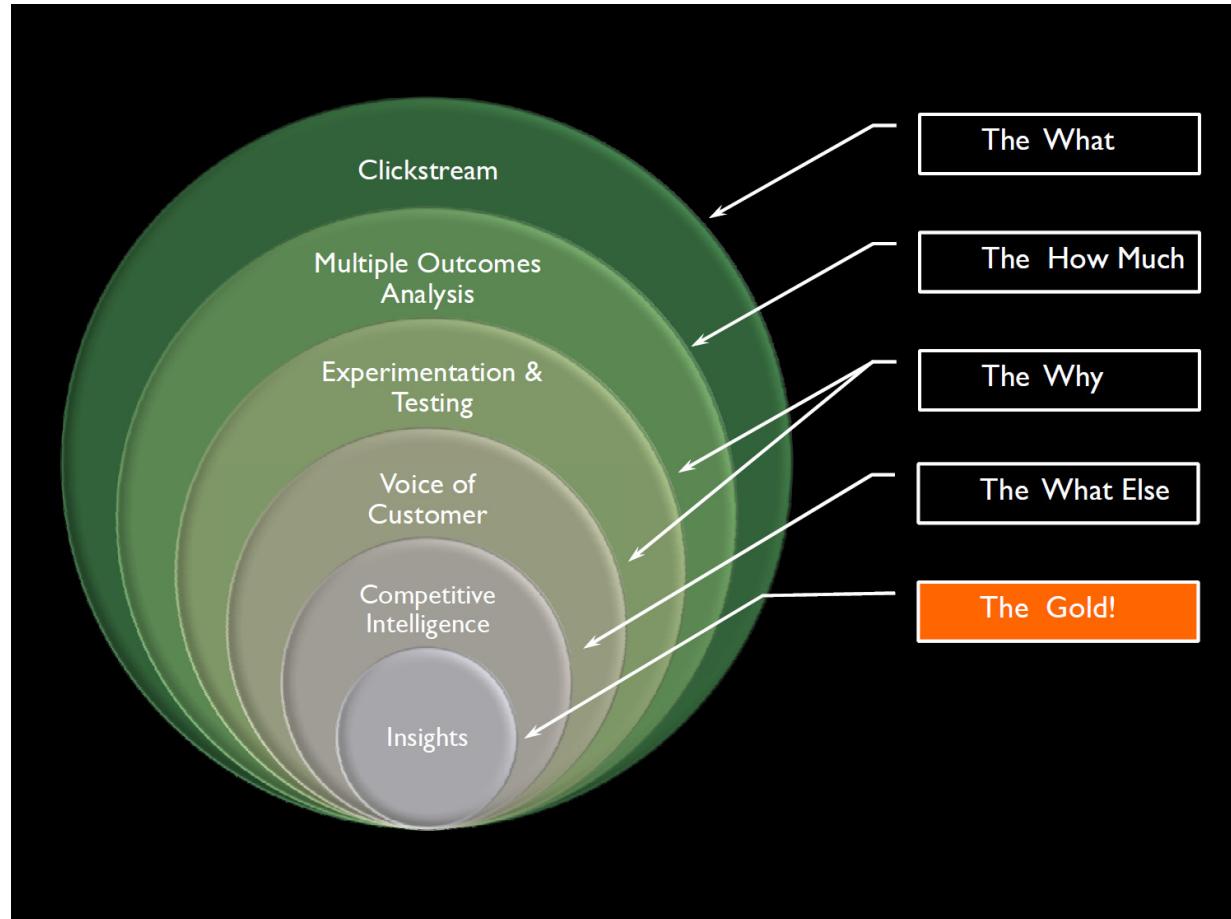
WEB ANALYTICS 2.0



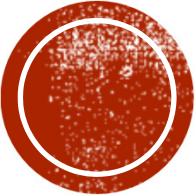
The What else?

- Do not ignore competition
- Compare your website trends with your competitors

WEB ANALYTICS 2.0



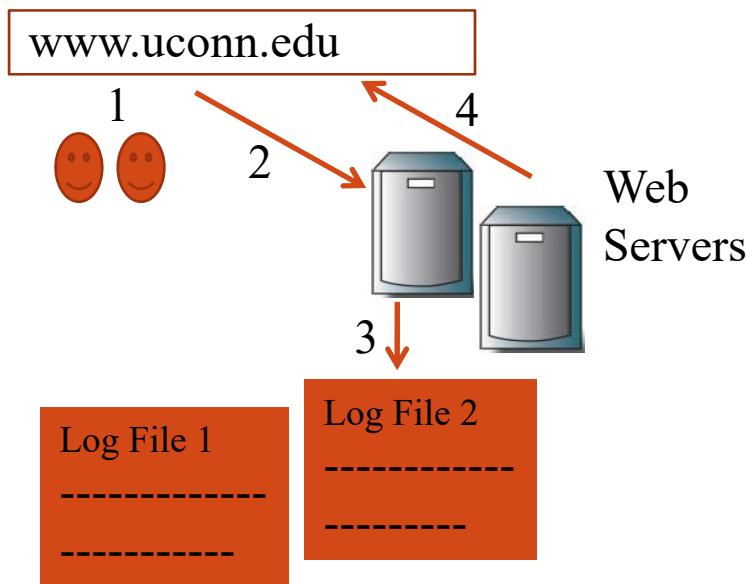
- Web Analytics 2.0**
is:
- (1) the analysis of **qualitative and quantitative data** from your website and the competition,
 - (2) to drive a **continual improvement** of the online experience that your customers and potential customers have,
 - (3) which translates into your **desired outcomes** (online and offline)



II. WEB DATA COLLECTION



DATA COLLECTION – WEB LOGS



▪ Process:

1. User requests webpage from browser.
2. Request for page comes to webserver
3. Web server accepts request and records entry in a log file (typical information logged includes date, time, browser, IP address etc.)
4. The webserver sends page to user

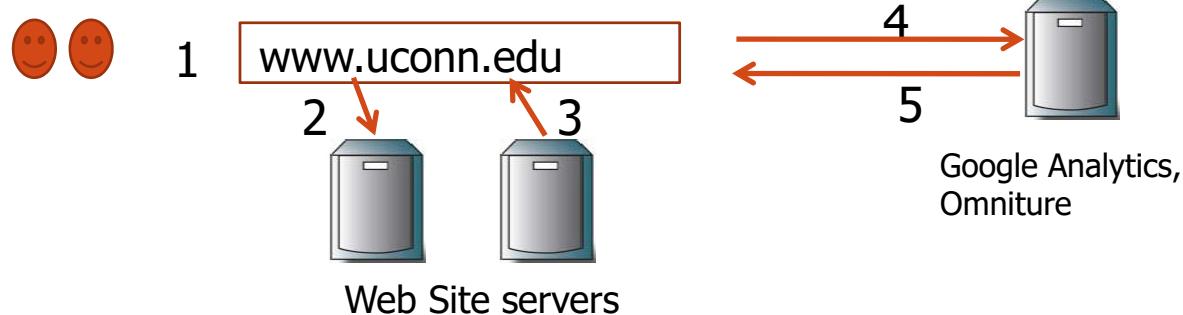
Pros:

- Easily Accessible
- Only mechanism that captures and stores visits of search engine robots
- You always own data

Cons:

- Require complex filtering
- Page caching – Some traffic becomes invisible

DATA COLLECTION – JAVASCRIPT TAGS



▪ Process:

1. User requests webpage from browser.
2. Request for page comes to webserver
3. The webserver sends page to user along with JavaScript code attached to page
4. As the page loads, the code gets executed. The code collects information related to page view, the visitor's session, and cookies. It then transmits that data to data-collection server.

Pros:

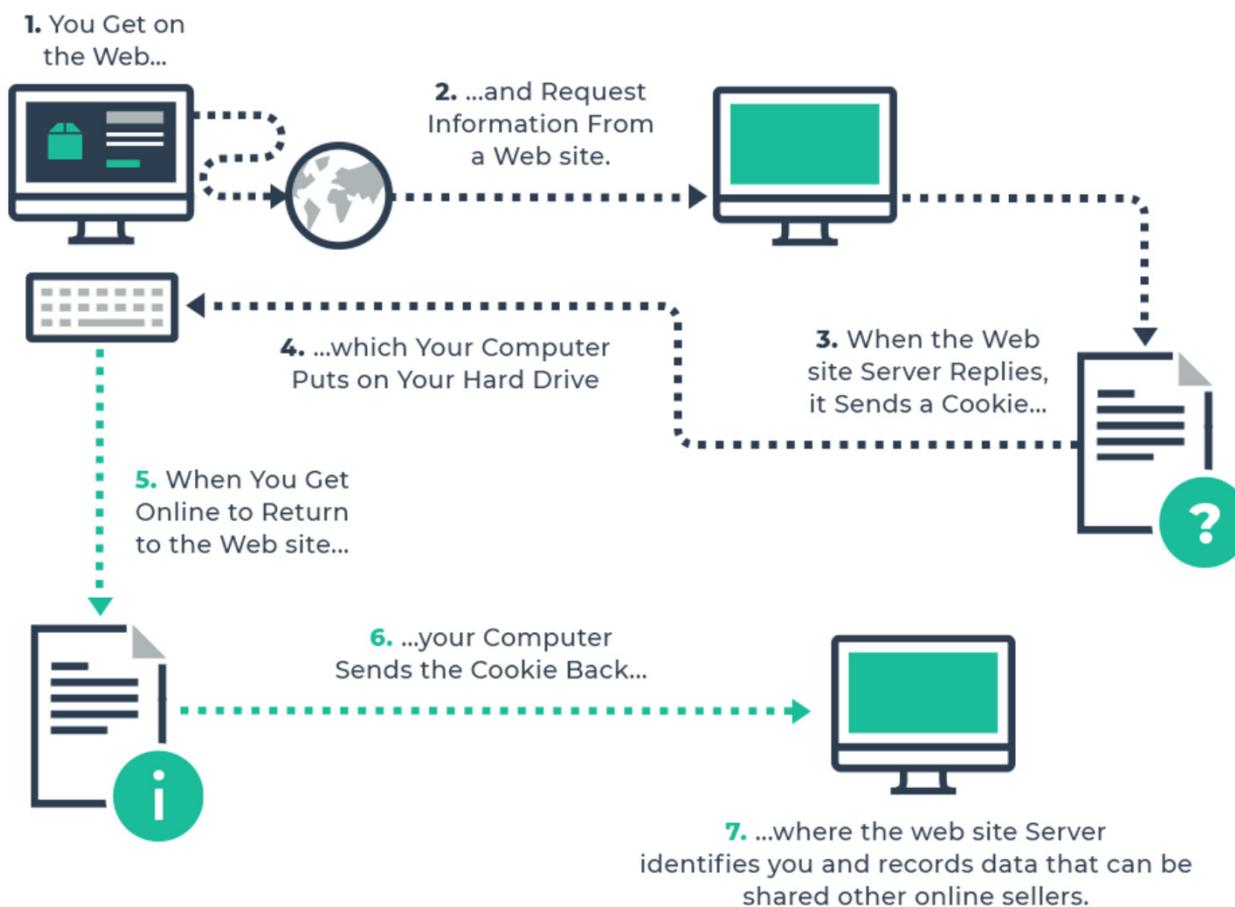
- Easy to implement and customizable
- Page caching does not affect data collection

Cons:

- 2-6 % visitors have JavaScript turned off

DATA COLLECTION – COOKIES

- Cookies are small text files stored locally, that are associated with visited website domains



DATA COLLECTION – COOKIES

- There are two types of cookies:

First Party Cookies

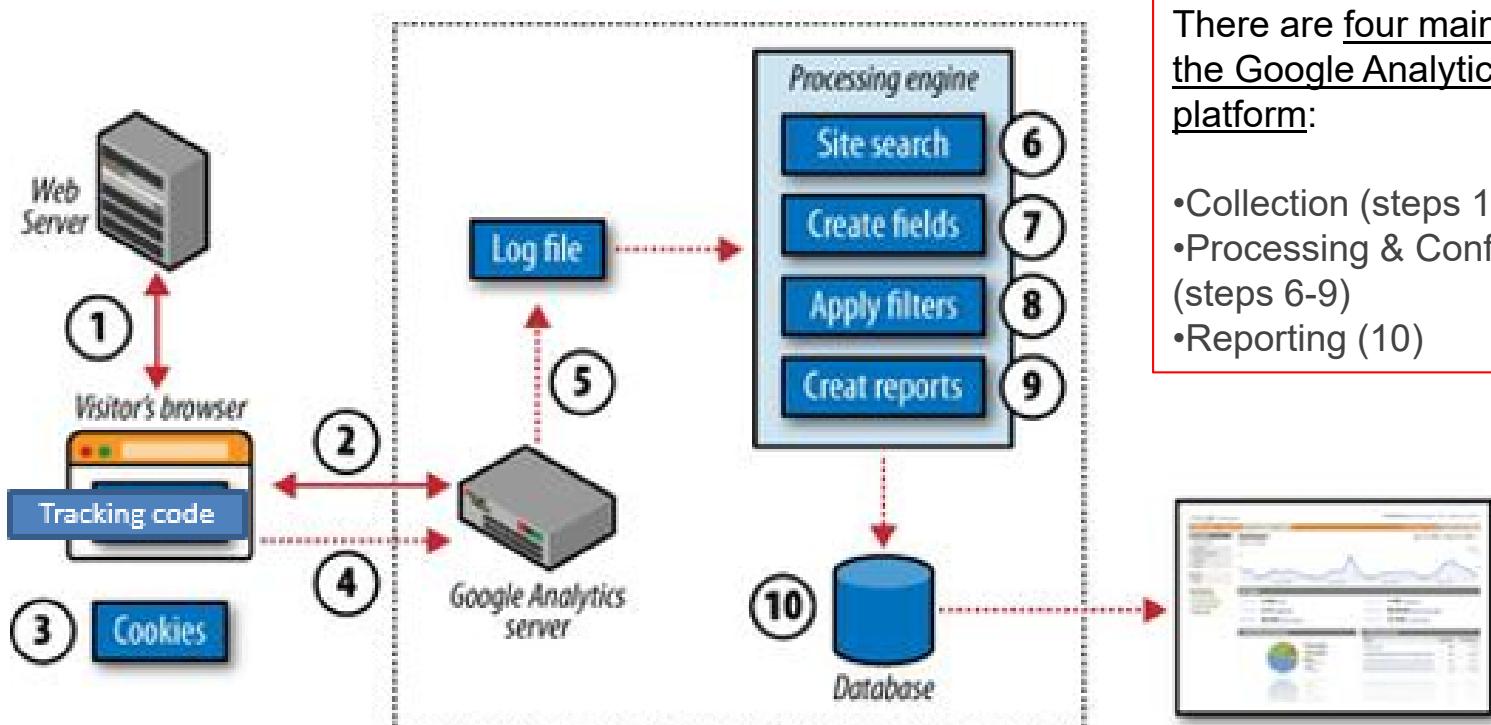
- A cookie that is created by website domain requested by the visitor.
- Only the website domain setting the cookie information can retrieve the data.

Third Party Cookies

- A cookie delivered by third party domain not directly requested by the visitor.
- The website domain setting the cookie can also list other domains allowed to view this information.

HOW GOOGLE ANALYTICS (GA) WORK

Google Analytics processing flow



- By default – GA uses Java script and first party cookie to collect data
- Cookie is used to identify unique visitors

HOW GOOGLE ANALYTICS (GA) WORK

- Role of cookies

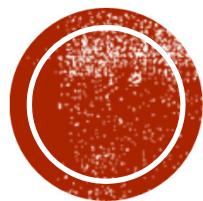
- GA cookie is set as a first party cookie because ~20% of users disable third party cookies
 - Cookies help to track customers across sessions

- Tracking code

- Basic: add a tracking code on every page of your website that you want to track
 - Advanced: use Google tag manager
 - When page loads on user's browser, the code executes and triggers a request to GA servers

TRACKING BEYOND COOKIES

- Browser fingerprinting –A combination of browser and OS characteristics are legitimately needed to properly display websites on a device, e.g.
 - Browser version
 - Screen dimensions
 - List of plugins
 - List of installed fonts
 - Language and time zone
- This combination is unique like a fingerprint, and can be used to identify a device
- Usage is quite widespread and major sites use them e.g. Skype, online banking sites etc.



III. TERMINOLOGIES



TERMINOLOGIES

- Business Objectives
- Goals
- Metrics
- Key Performance Indicators (KPIs)
- Targets
- Dimensions
- Segments

TERMINOLOGIES

- **Business Objectives** – "Why does your website exist?"

e.g., E-commerce sites: increase sales

Content sites: engage users

- **Goals** – Goals are specific strategies you'll leverage to accomplish your business objectives.

e.g., E-commerce sites: increase conversion, increase customer spend per order

Content sites: increase visits, increase user time on site

TERMINOLOGIES

- **Metric** is a quantifiable statistics that measures events or trends on a website
 - e.g., number of visitors, average time on site
- **Key Performance Indicator (KPI)** is a metric that helps you understand how well you are doing against your objectives, e.g.,
 - Conversion rate = # of conversions / # of visits
 - Average order value = total revenue / # of transactions
- **Target** is a pre-determined, numerical value as an indicator of success or failure
 - e.g., target average order value, benchmark from previous years (internal) or competitors (external)

KPI

- KPI should have the ability to provide recommendations for action which can significantly impact the business bottom line.
- Attributes of a good KPI:
 - Available and measurable
 - Highly impactful and relevant to its corresponding goal
 - Simple, instantly useful and timely
- How to find a good KPI?
 - Acquire a very good understanding of the business and its objectives.
 - Then translate the business objectives into measurable goals.
 - Once the goals are determined, select KPIs for each of these goals.

HOW WOULD THESE WEBSITES DEFINE THEIR OBJECTIVES?



The New York Times

WSJ

facebook

twitter

WHAT WOULD BE A GOOD KPI?

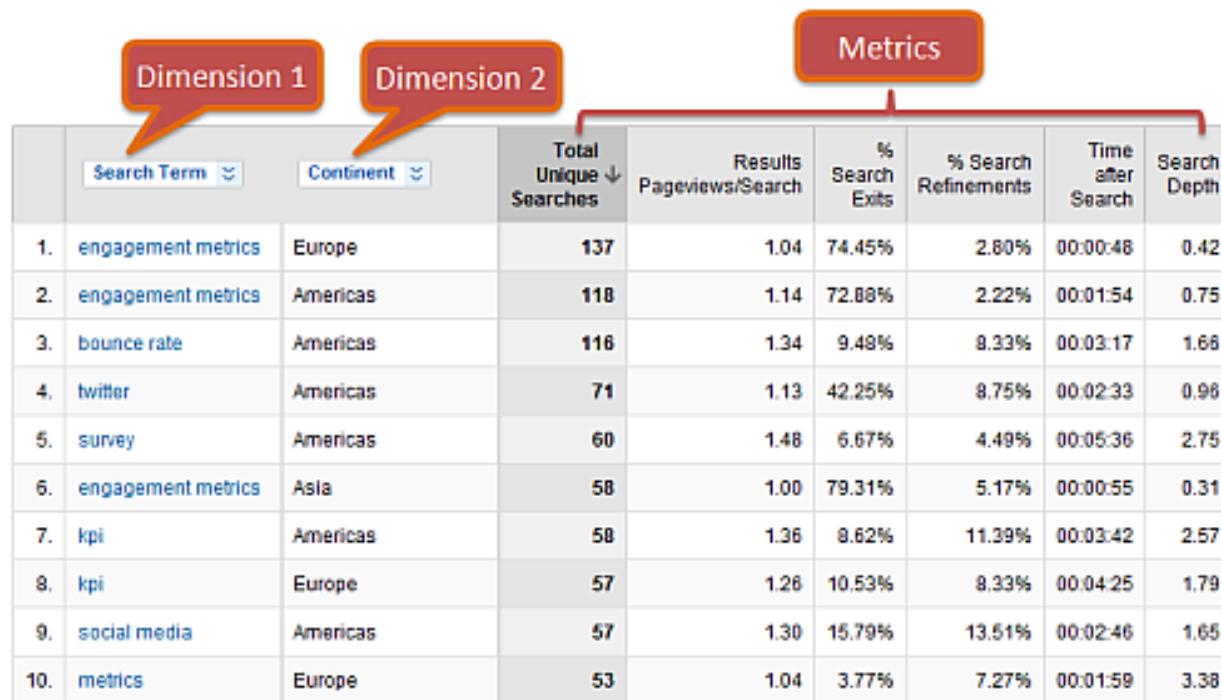
Company's Core Business Objectives	Business KPI
#1 Increase customer acquisition	
#2 Improve customer satisfaction	
#3 Increase sales	
#4 Decrease acquisition cost	

CONSIDER A TYPICAL RETAILER

- Is “number of Twitter followers” or “number of Facebook fans” a good KPI?

TERMINOLOGIES

- **Dimension** is an attribute or characteristics of an object
 - Dimensions allow you to group your data into different buckets to slice and dice the data.
- **Segment** is a group based on one or more **dimensions**
 - e.g., segment by date and time, device, marketing channel, geography, or customer characteristics (like repeat customers vs. first time customers) etc.



	Search Term	Continent	Total Unique Searches	Results Pageviews/Search	% Search Exits	% Search Refinements	Time after Search	Search Depth
1.	engagement metrics	Europe	137	1.04	74.45%	2.80%	00:00:48	0.42
2.	engagement metrics	Americas	118	1.14	72.88%	2.22%	00:01:54	0.75
3.	bounce rate	Americas	116	1.34	9.49%	8.33%	00:03:17	1.66
4.	twitter	Americas	71	1.13	42.25%	8.75%	00:02:33	0.96
5.	survey	Americas	60	1.48	6.67%	4.49%	00:05:36	2.75
6.	engagement metrics	Asia	58	1.00	79.31%	5.17%	00:00:55	0.31
7.	kpi	Americas	58	1.36	8.62%	11.39%	00:03:42	2.57
8.	kpi	Europe	57	1.26	10.53%	8.33%	00:04:25	1.79
9.	social media	Americas	57	1.30	15.79%	13.51%	00:02:46	1.66
10.	metrics	Europe	53	1.04	3.77%	7.27%	00:01:59	3.38

DIMENSION OR METRIC?

1. Browser
2. Type of mobile phone
3. Landing page name
4. Product purchased
5. Day of the visit
6. # of unique visitors
7. Average time on site
8. # of pages viewed
9. Average order size
10. Probability of abandoning the cart/checkout

SEGMENT BY TRAFFIC SOURCE

- Direct traffic
 - People who type url directly
 - Loyal customers, or offline marketing activities
- Referral traffic
 - From other websites who link to our website
 - Both paid and unpaid referrals possible (links on ads vs. links on blogs)
- Search traffic
 - From search engines
 - Organic results are unpaid traffic and sponsored links are paid traffic

EXAMPLE OF A RETAIL SITE

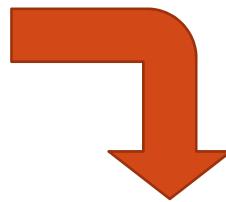
Visits	12,483
Conversions	512
Total revenue	\$17,541.12
Conversion rate	4.1%
Average order value	\$34.26
Revenue per visit	\$1.41

How useful is this information?

EXAMPLE OF A RETAIL SITE

Visits	12,483
Conversions	512
Total revenue	\$17,541.12
Conversion rate	4.1%
Average order value	\$34.26
Revenue per visit	\$1.41

How about now?



	Organic Search	Paid Search	Direct	Display Ads	Referral	Social Media	Email
Visits	2,219	1,430	141	2,238	4,140	792	1,613
Conversions	72	80	13	55	68	58	166
Total revenue	\$1,492.43	\$5,181.94	\$445.38	\$2,204.44	\$2,518.49	\$1,377.44	\$4,321.00
Conversion rate	3.4%	5.6%	9.2%	2.5%	1.6%	7.3%	10.3%
Average order value	\$20.73	\$64.77	\$34.26	\$40.08	\$37.04	\$23.75	\$26.03
Revenue per visit	\$0.70	\$3.62	\$3.16	\$0.99	\$0.61	\$1.74	\$2.68

SEGMENT BY LOCATION



Swissotel Hotels and Resorts are a chain of luxury hotels across the globe.

How did segmentation help?

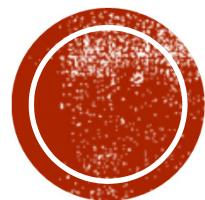
<https://www.digitalvidya.com/blog/swissotel-hotels-resorts-doubled-e-commerce-conversion-rate-by-adwords-optimization/>

THE STEPS

- Step 1: Identify the Business Objectives
- Step 2: Identify Goals for each Objective
- Step 3: Identify the Key Performance Indicators
- Step 4: Identify the Targets
- Step 5: Identify valuable Segments for analysis

AN EXAMPLE

Bike Company X Web Analytics Framework			
1 Business Objective	2 Goal	3 KPI	4 KPI Target
Sell Bike Parts	More Sales	Monthly Revenue	\$15,000 / mo
	Increase unique visits	Monthly Unique Visitors	13,000
	Make a Profit	Profit Margin/Sale	40%
Effective Marketing	CRM – build a customer DB	# of new registrations / mo	300 / mo
Build Goodwill	Draw Qualified Customers	Conversion Rate	3%
	Serve as resource to riding community	# of pageviews of resource pages	1500 / mo
5 Segmented KPI: (example) # of new registrations / mo			
Total Reg Goal = 300	Result = 332 (110% of Goal)		Percent
Paid Search	223		67%
Organic Search	67		20%
Referrals	17		5%
Direct	25		8%



IV. STANDARD METRICS



METRICS: PAGEVIEW & UNIQUE PAGEVIEW

- **Pageview** measures the number of times a page was viewed.
 - Individual activities on a web site send a request (hit) to the analytics servers. These include
 - Pageview (tracked by default)
 - Other events (e.g., click on an video or button) (need additional customization)
- In most analytics software including Google Analytics:
 - If a visitor hits reload after reaching the page, this will be counted as an additional pageview.
 - If a user navigates to a different page and then returns to the original page, an additional pageview will also be recorded.
- **Unique Pageview** metric will aggregates pageviews for the same page that are generated by the same user during the same visit/session
 - e.g., if a visitor views page A three times during one visit/session, it will be counted as three pageviews and one unique pageview

METRICS: UNIQUE VISITORS

- **Unique Visitors:** Count of all unique persistent cookie IDs during a given period of time.
- A visitor is uniquely identified by a cookie which assigns a random visitor ID to the user and combines it with the timestamp of the visitor's first visit.
 - Is each unique visitor a unique person?

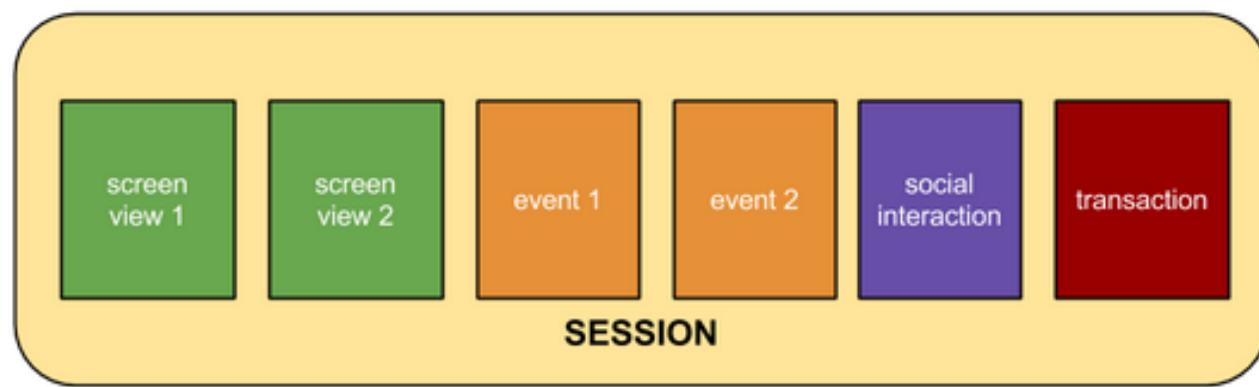
Example of counting unique visitors

- How to calculate the average daily unique visitors?

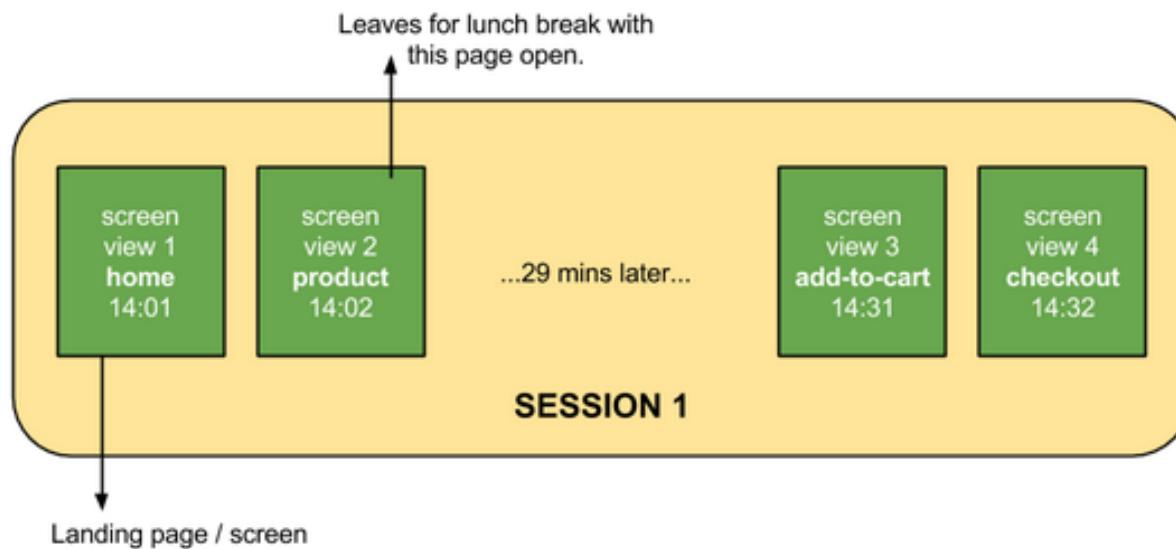
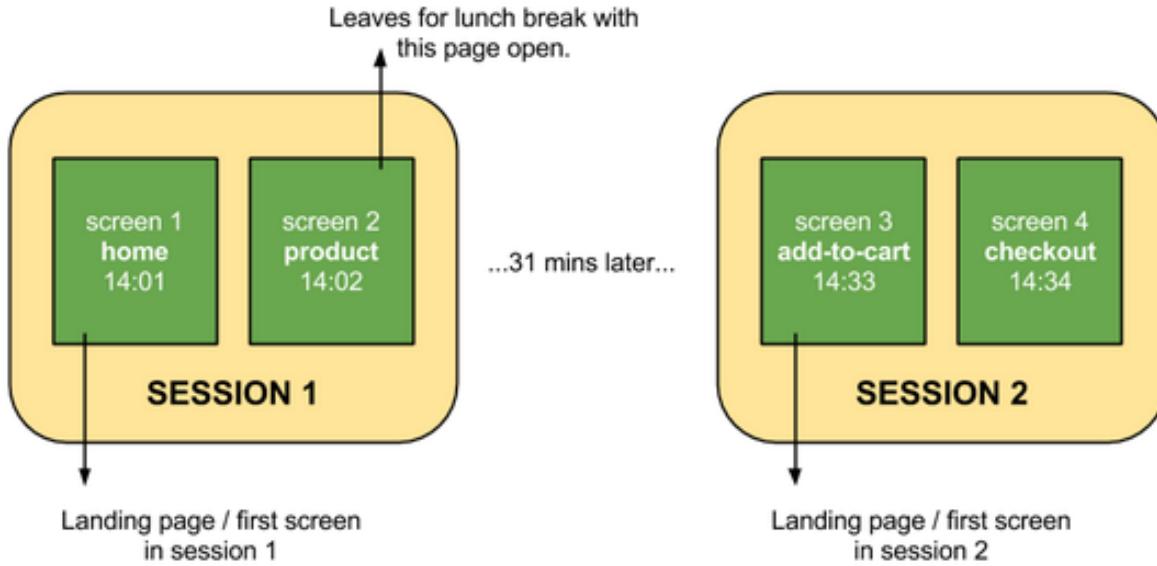
Month	Week	Day	Site Visits By			
1	1	1	Ann	Dennis	Matt	
1	1	2	Dennis	Matt	Dennis	
1	1	3	Matt	Matt		
1	2	1	Ian	Jim		
1	2	2	Jim	Ann	Bryan	
1	3	1	Jim	Ann	Bryan	
2	1	1	Jim	Ann	Bryan	Angle

METRICS: VISITS / SESSIONS

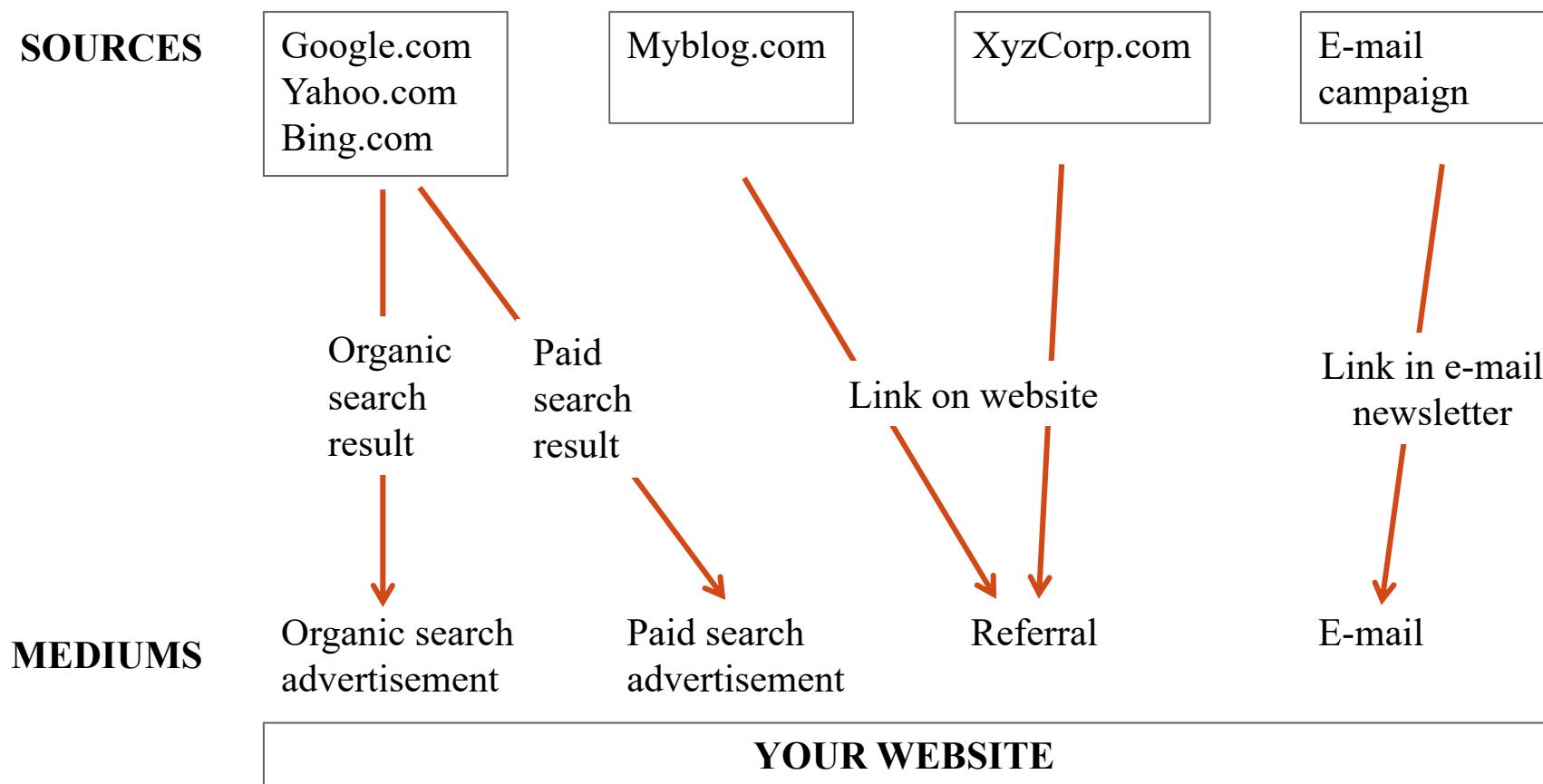
- A session is started when someone requests a page from your website.
- When a person leaves your website, a unique session ID is used to identify all the pages viewed in one cohesive visit.
- If the browser is left opened, session is terminated after fixed time of inactivity.
 - Google Analytics ends a session when:
 - more than 30 minutes have elapsed between pageviews by a single visitor.
 - campaign source value for the user changes.
 - at midnight.
 - If any of these events above occurs, then the next pageview from the user will be the start of a new session.



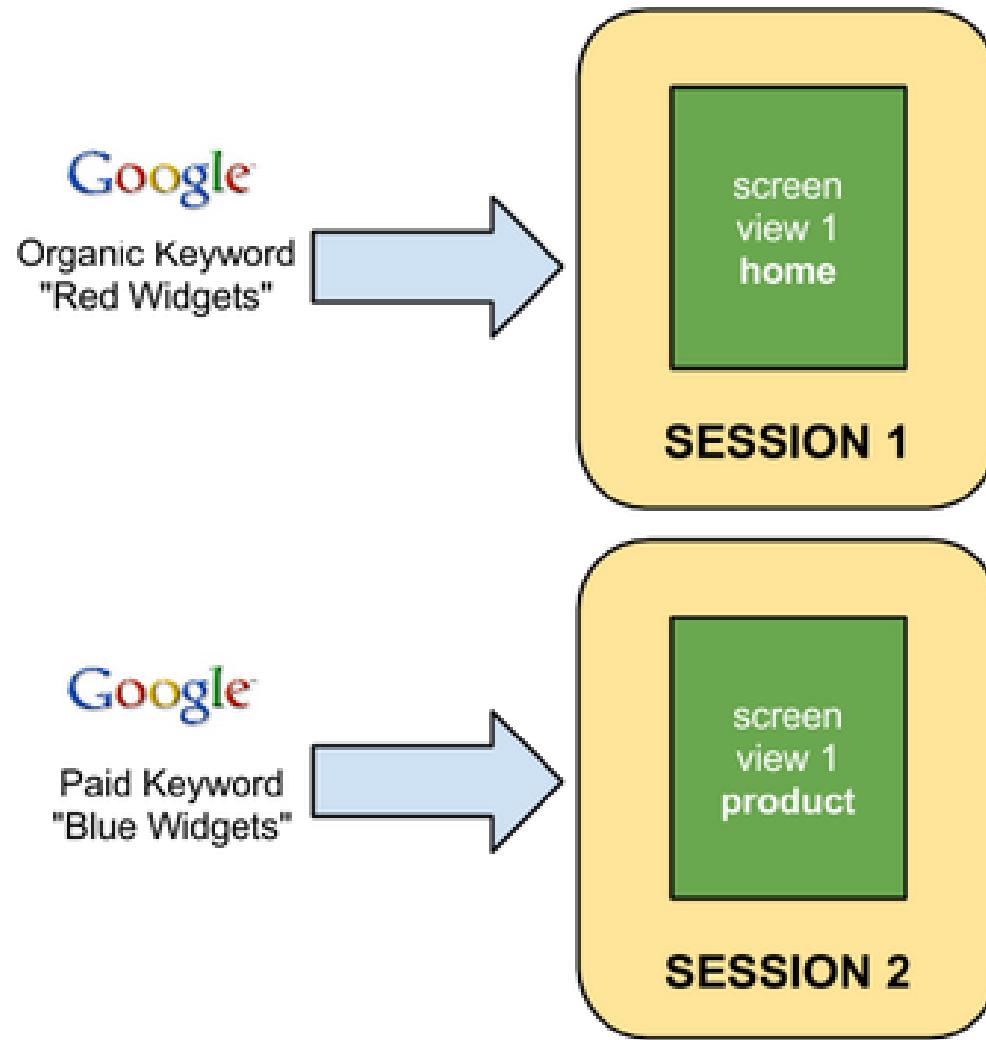
VISITS/SESSIONS BASED ON INACTIVITY



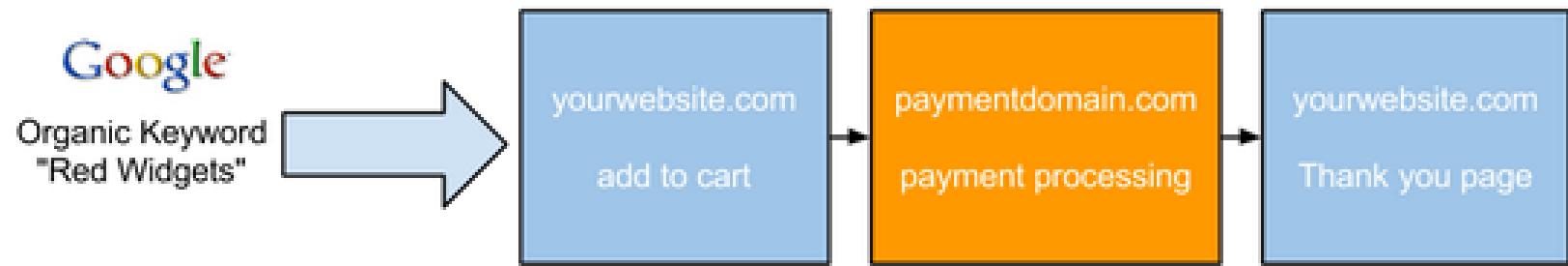
CAMPAIGN SOURCES



VISITS/SESSIONS BASED ON CAMPAIGN SOURCES

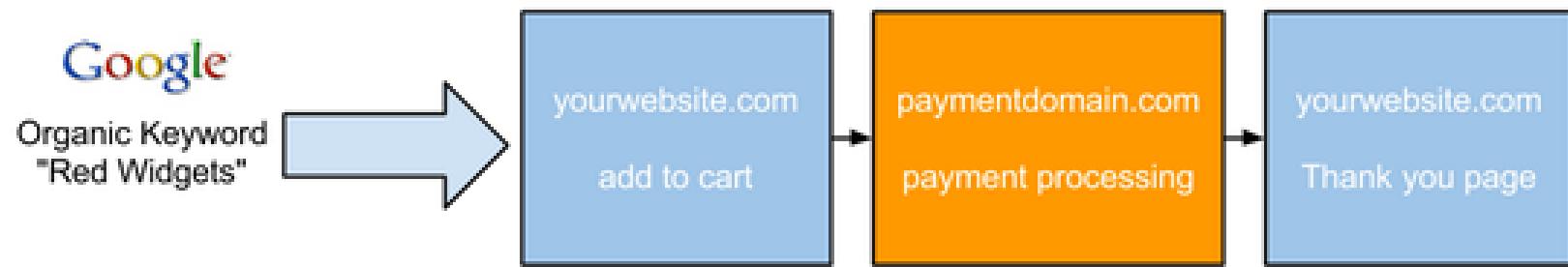


VISITS/SESSIONS BASED ON CAMPAIGN SOURCES



How many sessions ?

VISITS/SESSIONS BASED ON CAMPAIGN SOURCES

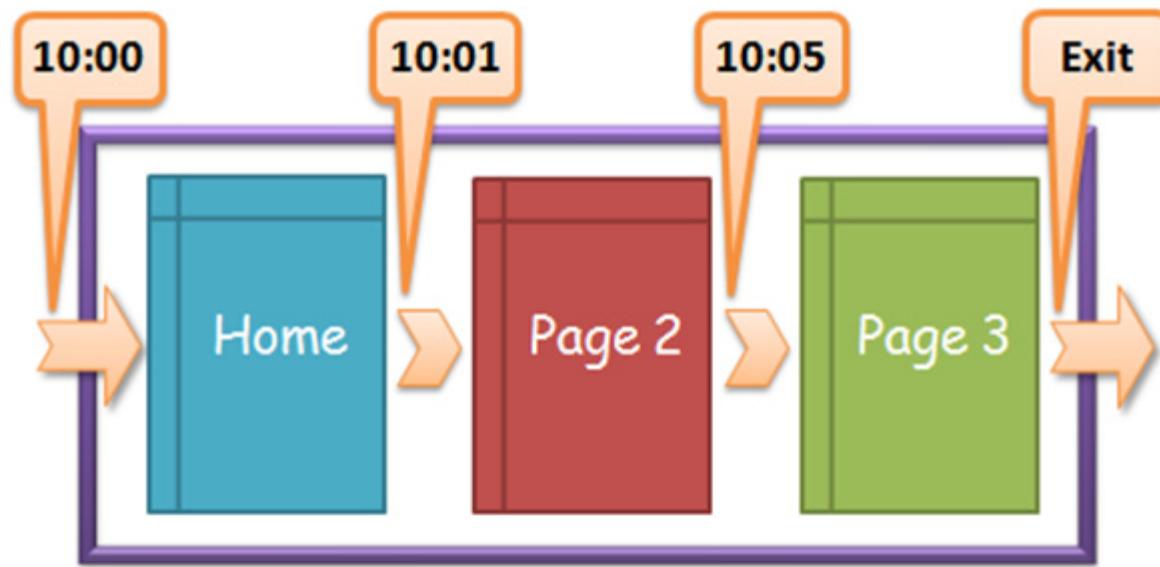


How many sessions ?

To consider this as one session – Add paymentdomain.com in
Referral exclusion list

METRICS: TIME ON PAGE & TIME ON SITE

- **Time on Page** measures the time spent on an individual page.
- **Time on Site** (also known as Session Duration) measures the time spent on the site during a visit (session).
- Note the missing time stamp on last page



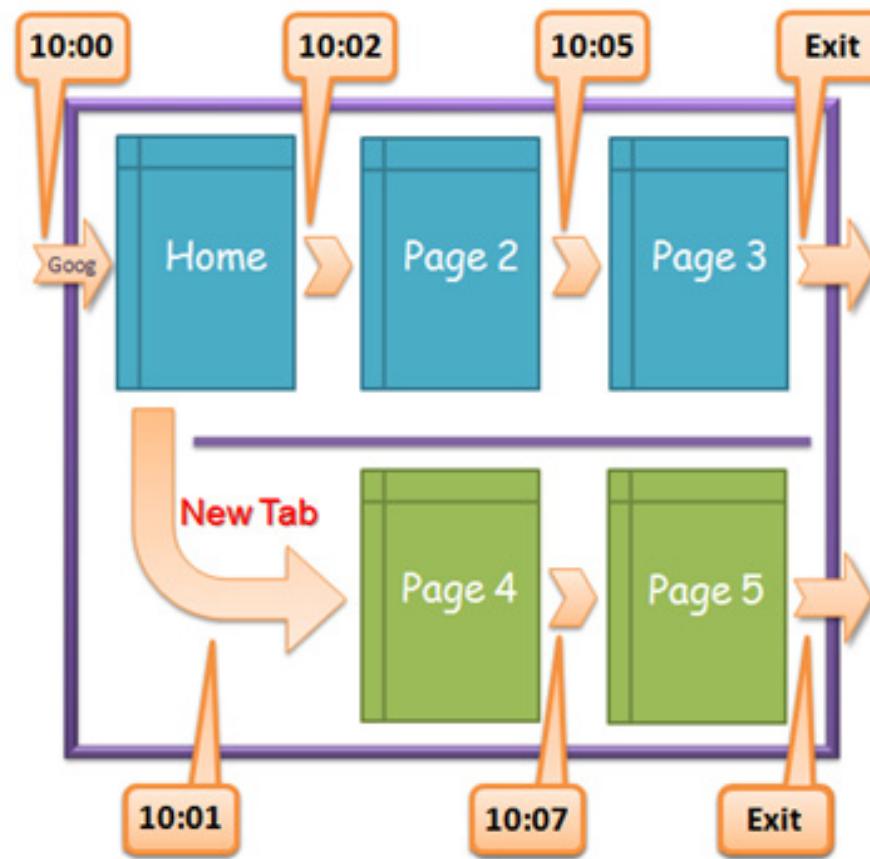
Time on Page (Home Page) = 1 minute.

Time on Page (Page 2) = 4 minutes.

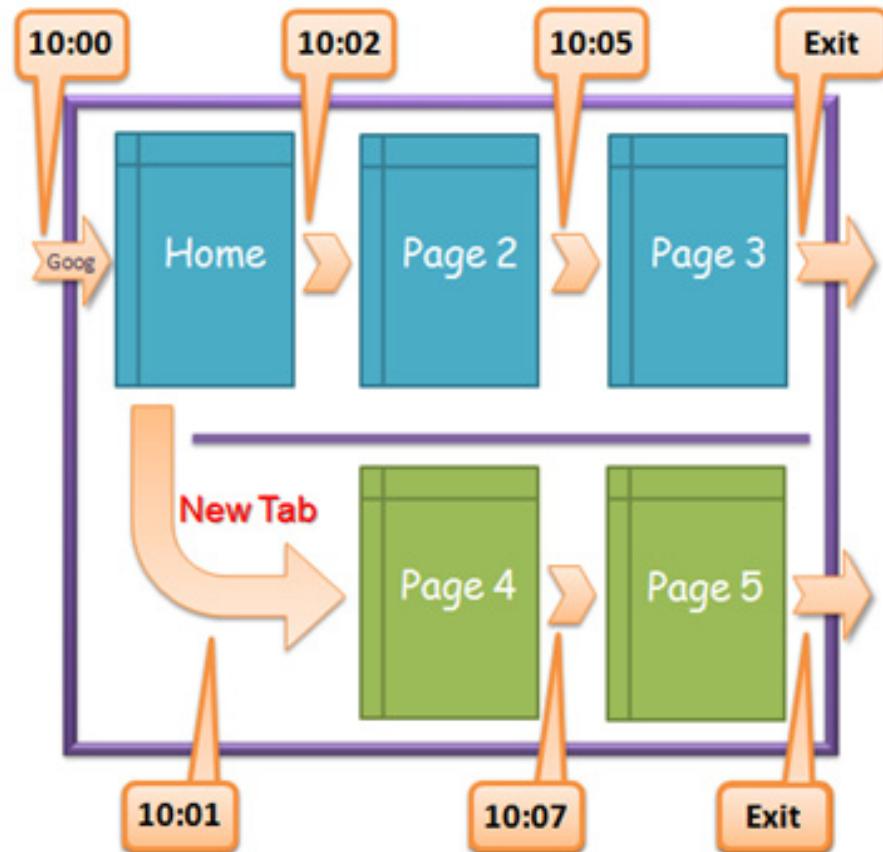
Time on Page (Page 3) = 0 minutes.

Time on Site = 5 minutes.

TIME ON PAGE & TIME ON SITE EXAMPLE



TIME ON PAGE & TIME ON SITE EXAMPLE: APPROACH #1



Two Sessions, one for each tab in the browser.

Session One (top): [referrer: Google]

Time on Page (Home Page) = 2 minutes

Time on Page (Page 2) = 3 minutes

Time on Page (Page 3) = 0 minutes

Time on Site = 5 minutes

Session Two (bottom): [referrer: your site/homepage]

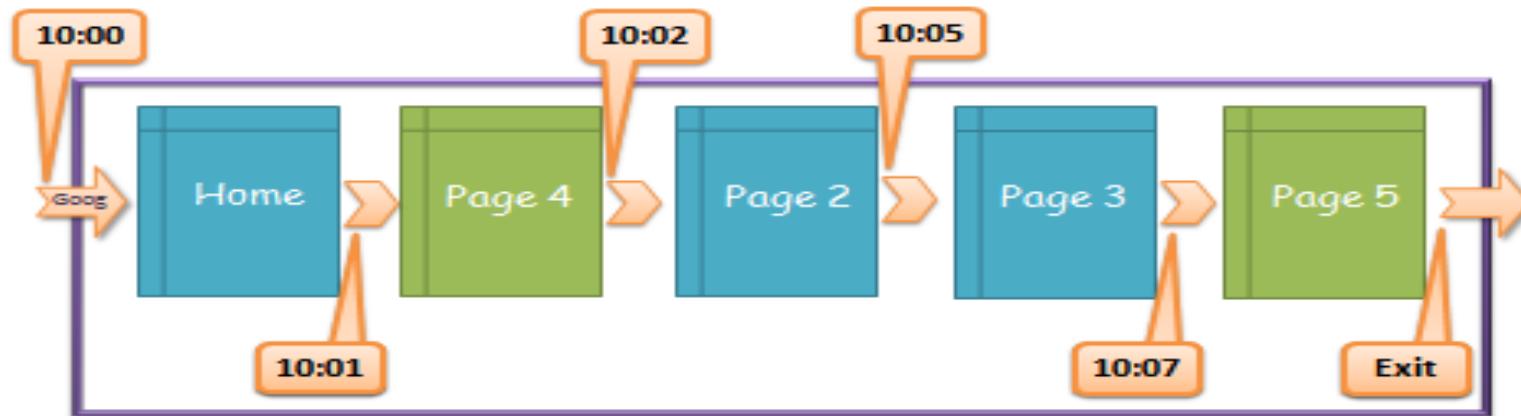
Time on Page (Page 4) = 6 minutes

Time on Page (Page 5) = 0 minutes

Time on Site = 6 minutes

2 Visits. 1 Unique Visitor.

TIME ON PAGE & TIME ON SITE EXAMPLE: APPROACH #2



One Session that combines the activities in the two tabs

Session One: [referrer: Google]

Time on Page (Home Page) = 1 minute

Time on Page (Page 4) = 1 minute

Time on Page (Page 2) = 3 minutes

Time on Page (Page 3) = 2 minutes

Time on Page (Page 5) = 0 minutes

Time on Site = 7 minutes

Google Analytics uses
approach #2

1 Visit. 1 Unique Visitor.

WHAT DOES THIS ALL MEAN?

- **Do not take the metric values literally**
- They are better suited for comparisons
- Context is also important
 - e.g., is Time on Site (or session duration) a good measure of user engagement?

ENTRANCE AND EXIT OF A SESSION

- An entrance occurs on the first page that a visitor accessed in a session.
 - This page is called the *landing page*.
- An exit occurs on the last page that a visitor accessed before a session ends.
 - This page is called the *exit page*.

METRICS: BOUNCE RATE

- **Bounce rate** is calculated as the percentage of sessions on your website with only one page view.
 - Measures the quality of traffic you are acquiring
 - Measures the percentage of people who come to your website and leave “instantly”
- Actionable on multiple levels
 - Aggregate level (entire site, regardless of the entry page)
 - Top landing pages (top entry page report): bounce rate for a page is based only on sessions that start with that page (i.e., the landing page), e.g.,

Page A → Page B → Page C

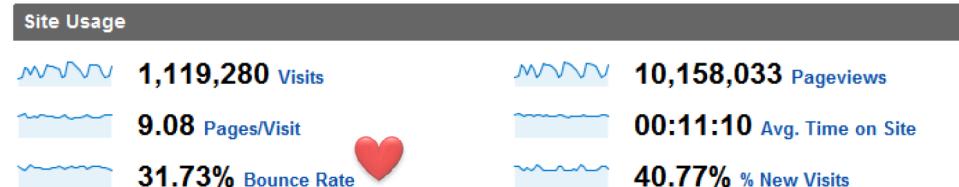
Page B → Page A → Page C

Page A → exit

Bounce rate for page A is **50%**.

BOUNCE RATE

- Can be used to evaluate
 - top landing pages
 - website referrers
 - keywords



- What could be the reasons for a high bounce rate?

BOUNCE RATE

- How are bounces included in calculation for average time on page and average time on site in Google Analytics?
 - Not included in average time on page, because the bounce has 0 time on page
 - Is included in average time on site (i.e., average session duration) – google analytics divides the total time for all sessions by # of sessions to calculate average session duration

E.g., 1000 sessions with 999 bounces. 1 session of 5 minutes.

$$\text{Average time on page} = 5/1 = 5 \text{ minutes}$$

$$\text{Average time on site} = 5/1000 = 0.3 \text{ seconds}$$

METRICS: EXIT RATE

- **Exit Page:** the last page on a site accessed during a visit, marking the end of a visit/session.
- **Exit Rate** =
$$\frac{\# \text{ of times a page is the exit page}}{\# \text{ of total pageviews of the page}}$$

— e.g.,

Visit 1: Page A → Page B → Page A → Page C → exit

Visit 2: Page B → Page A → Page C → exit

Visit 3: Page A → exit

Exit rate for Page A: **25%**

- Where users left can indicate what you need to fix

METRICS: CONVERSION RATE

- Conversion Rate =
$$\frac{\# \text{ of outcomes}}{\# \text{ of visits or unique visitors}}$$
 - Outcomes can be sales, course registration, forms filled, downloads, pageviews, etc.
 - What should be the denominator?
 - Use Visits if every visit can translate to conversion, e.g., sites where the same visitor will make multiple purchase within a short period of time.
 - Use Unique Visitors if a person visits website multiple times before making a purchase.
 - Google Analytics uses Visits by default.

OTHER METRICS

- Ad performance

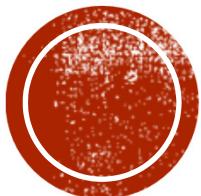
- Click through rate (CTR) =
$$\frac{\# \text{ of clicks on ad}}{\# \text{ of impressions (ad views)}}$$
 - Cost Per Impression (CPI) =
$$\frac{\text{total cost to an advertiser}}{\# \text{ of impressions}}$$
 - Cost Per Click (CPC) =
$$\frac{\text{total cost to an advertiser}}{\# \text{ of clicks}}$$

- Customer engagement

- # of new visits
 - # of repeat visits (frequency of visits)
 - Depth of visits (# of pages per visit)
 - Customer retention rate
 - Make repeat purchase (customer loyalty)
 - Frequently visit (customer recency)

- Great metrics are **simple, timely and instantly applicable**

- NOT all metrics are important to you: **find your own KPI**



V. WEB ANALYTICS STRATEGIES



VISITOR ACQUISITION

- Direct traffic
 - Less influenced by other factors
 - Great indicator for the overall health of the site
- Search engine
 - Organic search / paid search
- Social media, banner ads, emails, affiliates, etc.
- Referring sites
 - Blogs, forums, partners, competitors

OUTCOME BY TRAFFIC SOURCE



OUTCOME BY LANDING PAGE

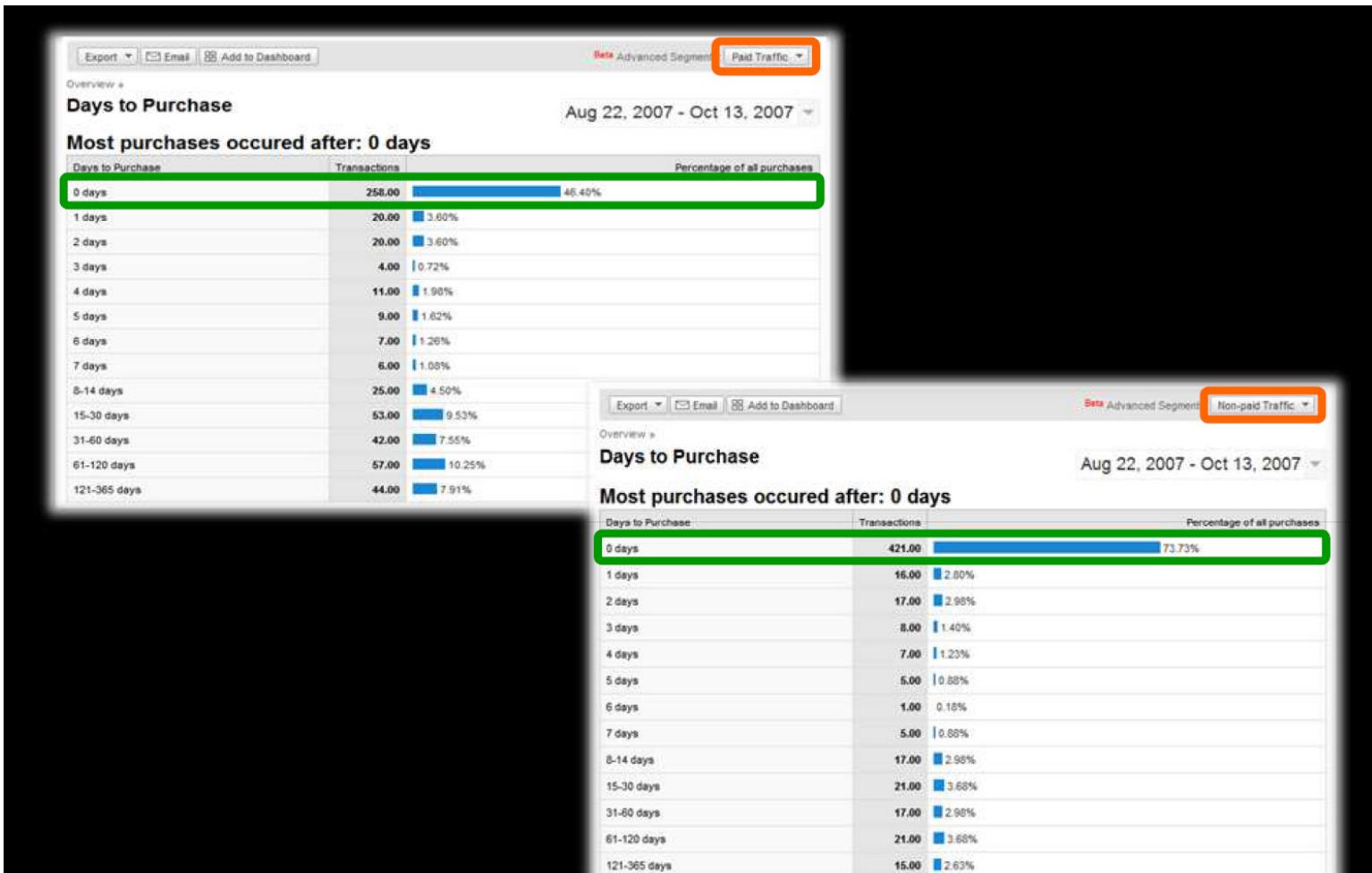
The screenshot shows the Google Analytics interface. On the left, a sidebar menu is open under 'Reports' with 'By URL' selected. A modal window titled 'REPORT PERIOD' is displayed, showing a calendar for the year 2008 with the range '2000-01-01 - 2000-12-31' selected. The main content area shows a report titled 'Custom www.kaushik.net/avinash'. The reporting period is set to '2008-01-01 - 2008-12-31 (PST)'. The report table lists various URLs with their visit counts, page views, and bounce rates. Several bounce rates are highlighted with orange boxes: 87.40%, 82.76%, 94.47%, and 71.13%.

Entry Page URL	Visits	Page Views	Bounce rate
http://www.kaushik.net/avinash/	41,000	82,185	64.74%
http://www.../10-insights-from-11-months-of-working-at-google.html	16,708	20,559	87.40%
http://www.kaushik.../the-three-greatest-survey-questions-ever.html	8,311	9,935	90.11%
http://www.kaushik.net/avinash/20.../web-analytics-demystified.html	6,317	11,835	66.88%
...e-analytics-is-re-launched-do-these-five-things-first-in-v2.html	4,327	5,815	82.76%
http://w.../4q-the-best-online-survey-for-a-website-yours-free.html	3,774	5,427	79.28%
...blog-metrics-six-recommendations-for-measuring-your-success.html	3,386	4,694	81.16%
...as-customer-value-customer-retention-and-non-line-marketing.html	2,837	3,040	94.47%
...nt-analytics-lip-11-measure-effectiveness-of-your-web-pages.html	2,650	4,668	71.13%
ht.../standard-metrics-revisited-time-on-page-and-time-on-site.html	2,502	4,107	75.54%
Subtotal	91,012	152,265	74.76%
Total	176,838	291,465	74.56%

« PREVIOUS 10 NEXT 10 »

Which *home pages* are your best friends?

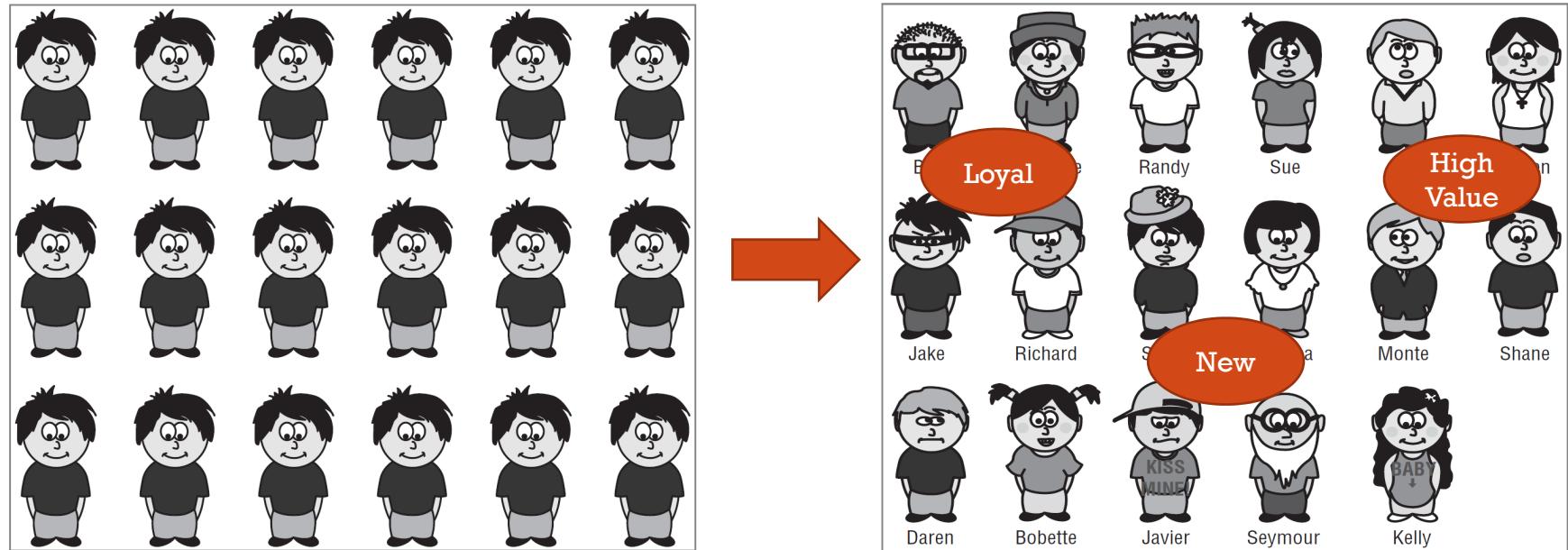
OUTCOME BY PAID VS ORGANIC



Paid vs. Organic Purchase Behavior – Understand & Optimize.

SEGMENTATION: ONE SIZE DOES NOT FIT ALL

- Segmentation gives you insights into different visitors' behavior, different channels' performance, site visitor loyalty (new vs. returning visitors) and enables you to develop customized marketing initiatives and tailored user experiences for these segments.



CONTEXT, CONTEXT, CONTEXT

- Comparing key metrics performance for different time periods
- Segmenting
- Comparing key metrics and segments against site average
- Pair/group key metrics together
 - E.g., visits & percentage of new visitors
- Leveraging industry benchmarks and competitive index
- Leveraging tribal knowledge
 - initiatives, management changes, marketing programs, website updates, marketing campaigns, etc.

EXERCISES

1. Calculate exit rate and bounce rate for Page A, B and C

visit 1: B → A → C → exit

visit 2: B → exit

visit 3: A → B → C → B → exit

visit 4: C → exit

visit 5: B → C → A → exit

2. If someone comes to your site and views page A, then page B, then Page A again, and then leaves your site, the total pageviews for the visit is ____, and the unique pageviews for the visit is ____.

- A. 1
- B. 2
- C. 3
- D. 4

3. A media site with lots of written articles should probably use _____ as a KPI, whereas a media site with mostly video content should use _____ as a KPI.
4. 1,000 visits to a site generated \$500 in revenue selling product that cost \$250. How much can you set as CPC at most?