



**MET CS688 C1**

# ***WEB ANALYTICS AND MINING***

**ZLATKO VASILKOSKI**

WEB ANALYTICS – BASIC TERMS

# Web Analytics Overview

- Web analytics goal:
  - Improving the online experience of your customers and potential customers by data analysis from your business (and the competition).
  - Many examples you might be familiar with (real estate, pharmaceuticals, travel etc.). [Can you suggest few more?](#)
- Analytics program includes aspects of
  - Collecting relevant raw data
  - Understand significance contained in the data
- The focus is to understand the interaction with the customer
  - How the search keywords (and advertisements) influence the search process so that the business can have more visitors (potential customers).
  - Understand (measure) the user experience, behavior and satisfaction with the web site.
- Needed to achieve this
  - Infrastructure to collect and process data (technology)
  - Skills to analyze and interpret that data (qualified people)
- Common misunderstanding of comprehensive web analytics program
  - Focusing on collecting and reporting raw data without understanding their significance
- Skills needed: understanding statistics, mathematics.
  - Recently attempts to create tools to compensate for these skills.
- Fluidic and evolving subject in nature, roles are still being defined
  - Who is responsible for web analytics? Marketing or IT?

# Web Analytics Overview

- Goal - Illustrate how to establish a comprehensive analytics program.
- Good tools are needed for any implementation of analytics.
- Different tools are needed for different types of businesses.

# Example 1

- To better target, acquire, and retain customers, marketers need to use data analytics, content marketing, and customer engagement.
- For example an office furniture retailer could increase its return on what it spends for advertising by use of data analytics in the following way.
- To acquire new customers, the retailer has to figure out better ways to find potential clients to target for its ads. To do that, however, the company needs huge amounts of **intent**—or **in-market**—data.
- **Intent data** is data collected about online users' activities—indicating some future action, or intent, such as ordering a product. This can be achieved by obtaining web site's key performance indicators using web (or free Google) analytics as described in the first part of this semester class notes.
- When potential customers is looking to furnish a new office may interact with the retailer's web site in a variety of ways, such as browse through architectural sites for design ideas, visit various office retail sites to evaluate items or perhaps even do some comparative analysis on a product review site. All these actions signal that the consumer is actively browsing, researching or comparing the types of products online furniture retailers sell. Customer's intents, preferences, and loyalties create impressions and Web/Google analytics tools can then capture that data enabling marketers to act on it.

# Example 2

- ENERGY STAR, a score from 0 to 100, is a measure of energy consumption performance. A score of 40 means performance worse than 40% of similar buildings nationwide. A score of 75 or higher makes you eligible for ENERGY STAR certification.
- Energy Management Associates, Inc. Help customers achieve ENERGY STAR certification rating by reviewing customer's portfolio, and providing calls with questions or suggestions, to raise customer's score and provide up to 20% savings opportunities.
- According to EMA's web site, customers are encouraged to provide a copy of their most recent utility invoice (12 to 14 month usage history) for each account (gas and electric). This data is contained in EMA's database that with help of **predictive data analytics**, can be used to enhance the detailed customer profile and predict the most likely factors affecting the ENERGY STAR score, by integrating an automatic analysis of the key performance indicators with historical data, by region, type of energy etc. and tied together with external data available on the web, such as location specific temperature and weather conditions.

# Example 3

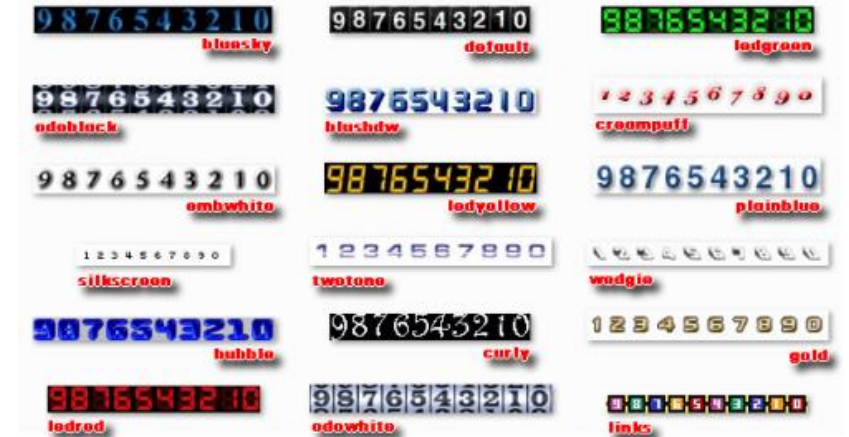
- The use of **predictive data analytics** enables the retailer, insurance company, the travel agency or any similar eCommerce (click to order) business to predict the most likely purchase type from a given customer.
- This is typically achieved by leveraging historical data (sometimes tied together with a third-party data) to paint a detailed picture of the buyer profile. From that information, it can be determined:
  - Which customer will purchase
  - What product they will purchase
  - What message they will respond to
  - Which customers will focus on
- Predictive data analytics can recognize patterns and behaviors for more effective
  - messaging
  - plan persona-specific campaigns based on customer's habits and past preferences.

# Example 4

- Data analytics is also commonly used to:
  - Give real-time visibility of inventory and warehouse management systems.
  - Sifting through an abundance of social media information. Monitoring, analyzing and reporting on the voice of its customers on social media including Facebook, Twitter, and various blogs and forums.
  - To more accurately analyze the voice of the customer, predictive analytics (and NLP - Natural Language Processing) can be used to score data attributes and determine which social media posts are actionable and relevant thus filtering out the noise of irrelevant posts.

# Brief History of Web Analytics

- Every time someone locates a website the web server logs data.
- Before 1995, simple reports based on information that is automatically collected
  - filename
  - time, referrer (i.e., the website forwarding the request)
  - browser
  - operating system
  - computer data
- Later hit counter introduced on websites, but not accurate so they vanished.
- WebTrends usually associated with first commercial web analytics programs.





# Brief History of Web Analytics

- Several years later, introduced website analytics software that was able to measure click density or site overlay and heat maps.
  - Enabled to understand exactly, which links and where on a page, visitors were clicking.
  - Reports could described the consumer behavior by reporting number of clicks or overall percentage for the identified web pages.
- Large amount of data collected by vendors from every web site visit and click – **clickstream**.
  - It describes what data is collected.
  - It does not contain any insight into the significance of the data.
  - Too large to be useful.
- Big web analytics vendors:
  - Google, WebTrends, Coremetrics, Omniture, WebSideStory, IBM and Adobe.
  - Please have a look at SiteCatalyst a tool from Omniture (Adobe marketing cloud) to see the features and functionalities that can be found in modern tools.
- Ever changing and growing industry, gained much more popularity recently with Big Data.

# Addressing the “What” question

Important web analytics questions to address - “What happened?”

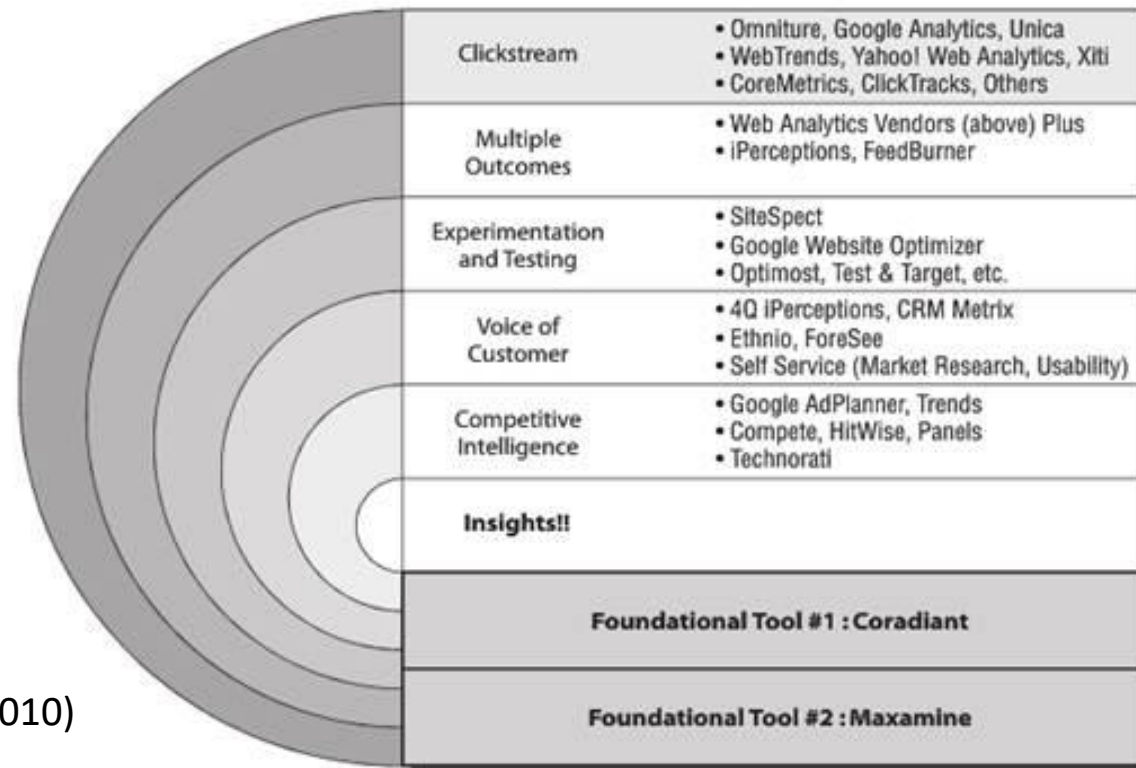
- What pages did people view on your website?
- What products did people purchase?
- What was the average time spent?
- What sources did they come from?
- What keywords and campaigns produced the clicks?

# Addressing the “Why” question

- Even more critical is to know “why” people do the things they do on your web site
  - This can be used to make intelligent decisions about your web presence.
- Important “why” questions to address
  - Analyze qualitative and quantitative data from our web site and a competitor’s web site
  - Focus on continual improvement of the customer’s online experience
  - Better translate data into desired outcomes, both on- and off- line
  - Decide whether we should go with channels like TV and radio advertising, rather than online advertising
  - Obtain a strategic advantage over competitors who focus only on clickstream
- Examples:
  - dropped shopping carts or
  - dropped registrations forms

# Web Analytics Tools

- Which web analytics tools to use?
- Depends on the size of your business and the available resources
  - Small Business – Click stream, outcomes, voice of the customer
  - Medium Business – All of the above plus Testing
  - Large Business – All of the above plus Competitive intelligence, Deep back-end analysis, Site structure gaps



From “Strategy and Tools”  
Webanalytics 2.0 by Kaushik (Wiley 2010)

# Goals, Context,

- It is Important to know
  - Goals of Analysis
    - Target that lets us measure success
    - Assumes understanding both the “what and the why”
    - Examples:
      - Knowing where visitors come from.
      - Understanding what they did in the website.
      - Did they meet the goal?
      - Did the consumers respond to the online marketing plan in the manner you expected?
      - Does the data analytics tool validate that?
  - Context Study
    - Website visitors are made up of groups of people behave differently and have different objectives.
    - Their action or behavior can be analyzed with the web analytics tools.

# and Segments

- Segment (determine as much as possible features for classification)
  - Analyze particular groups (segments) of people that come to the website.
  - What this particular group has in common with other groups.
  - Relate visitors by geographical regions (east/west coast, Europe, Asia etc.)
  - Relate visitors by gender and their preferences.
  - Segment data by visitor's motivation.
  - Which group of people “browsed, shopped and purchased”.
  - Was this the group that you sent out an email marketing outreach newsletter?
  - If so do we need to do more outreach via e-newsletters.

# Defining Basic Analytics Metrics

- In analytics it is essential to know which numbers are important and why.
- Note that the metric focuses on “Why”.
- Basic Analytics Metrics
  - Visits and Visitor Sessions
  - Referrals
  - Bounce & Exit Rate
  - Conversion Rate
  - Engagement
  - SEO, Social Media, Emails and Metrics

# Visits and Visitor Sessions

- **Visitor** - an individual (not necessarily a human) or device such as browser which accesses a Web site within a specific time period.
  - Unique visitor within a specific reporting period (no double counting).
- **Visit (Sessions)** – an interaction with a data source (examp: text and/or graphics downloads) from a single browser (device) during a single session.
  - A visit can consist of a series of page views that a single visitor makes during a period of browsing activity. A visit ends after the visitor closes the browser, clears cookies, or is inactive for 30 minutes (customizable time period).
  - During each visit, users will engage in one or more interactions with the web site pages.
  - Analytics software will automatically track these interactions as “**pageviews**.” The pageview metric increases every time a page is viewed on your site. Other activity, like watching a video, mouse position, etc. can also be tracked. Such activities are better classified as “**events**” rather than pageviews.
  - **Cookie** (persistent or session) is a file on the user’s device that identifies the user’s unique browser.
  - Tracking code looks for cookies. If a cookie is deleted or blocked incorrectly counts unique visitors.



# Referrals - Where do visitors come from?

- **Referrals** indicate the place from which the user clicked to get to the current page.
- It is valuable to know how someone found your web site. Was it
  - through a search engine
  - positive review
  - social-media talk
  - email or e-newsletters
- Referrals are the lifeline for marketing advertisements.
- It's important to know which campaigns helped draw in new visitors or succeeded in getting loyal customers.

# Bounce and Exit Rate

- **Bounce rate** and the **exit rate** measure whether users find a web site or a web page useful.
- Bounces are counted for users who land on a page and leave immediately. They do not see the page content.
- Reasons can be site-design or usability issues or many other reasons.
- Typically expressed as percentage of single-page sessions.
- If the exit rate is high, the exit-rate metric can be meaningless, and it should not matter.
- Useful to find out if visitors are abandoning the site at a certain point in the middle of an e-commerce transaction.

# Bounce and Exit Rate

- **Bounce rate** and the **exit rate** measure whether users find a web site or a web page useful.
  - Bounce Rate: the percentage of single-engagement sessions. Determined by calculating the number of bounces over the total number of pageviews to a page.
  - Exit Rate: the percentage of exits on a page
    - **Monday:** Page A > Page B > Page C
    - **Tuesday:** Page B > Page A > Page C
    - **Wednesday:** Page A > exit
  - Report for Page A would show 3 pageviews and a 50% bounce rate since Tuesday's pageview did not start with Page A.

## Bounce Rate

$$R_b = \frac{T_v}{T_e}$$

$R_b$  = Bounce Rate  
 $T_v$  = Total one-page visits  
 $T_e$  = Total entrance visits

## Exit Rate

$$R_e = \frac{T_{\text{exits}}}{T_{\text{views}}}$$

$R_e$  = Exit Rate  
 $T_{\text{exits}}$  = Total exits from page  
 $T_{\text{views}}$  = Total visits to page

- Reference: <https://support.google.com/analytics/answer/2525491?hl=en>

# Bounce and Exit Rate

What are the Bounce and the exit rate considering the following single-session days

- **Monday:** Page B > Page A > Page C > Exit
- **Tuesday:** Page B > Exit
- **Wednesday:** Page A > Page C > Page B > Exit
- **Thursday:** Page C > Exit
- **Friday:** Page B > Page C > Page A > Exit

## Bounce Rate

$$R_b = \frac{T_v}{T_e}$$

$R_b$  = Bounce Rate  
 $T_v$  = Total one-page visits  
 $T_e$  = Total entrance visits

## Exit Rate

$$R_e = \frac{T_{\text{exits}}}{T_{\text{views}}}$$

$R_e$  = Exit Rate  
 $T_{\text{exits}}$  = Total exits from page  
 $T_{\text{views}}$  = Total visits to page

# Bounce and Exit Rate

What are the Bounce and the exit rate considering the following single-session days

- **Monday:** Page B > Page A > Page C > Exit
- **Tuesday:** Page B > Exit
- **Wednesday:** Page A > Page C > Page B > Exit
- **Thursday:** Page C > Exit
- **Friday:** Page B > Page C > Page A > Exit

*Bounce Rate:*

- Page A: 0% (one session began with Page A, but that was not a single-page session, so it has no *Bounce Rate*)
- Page B: 33% (*Bounce Rate* is less than *Exit Rate*, because 3 sessions started with Page B, with one leading to a bounce)
- Page C: 100% (one session started with Page C, and it lead to a bounce)

*Exit Rate:*

- Page A: 33% (3 sessions included Page A, 1 session exited from Page A)
- Page B: 50% (4 sessions included Page B, 2 sessions exited from Page B)
- Page C: 50% (4 sessions included Page C, 2 sessions exited from Page C)

## Bounce Rate

$$R_b = \frac{T_v}{T_e}$$

$R_b$  = Bounce Rate  
 $T_v$  = Total one-page visits  
 $T_e$  = Total entrance visits

## Exit Rate

$$R_e = \frac{T_{\text{exits}}}{T_{\text{views}}}$$

$R_e$  = Exit Rate  
 $T_{\text{exits}}$  = Total exits from page  
 $T_{\text{views}}$  = Total visits to page

# Conversion Rate

- In the context of studying goals and outcomes, this metrics is a significant one.
- The **conversion rate** (as a percentage), is defined as
$$\text{conversion rate} = \text{outcome} / \text{unique visits} * 100$$
- An example of an **outcome** could be something very simple like clicking on ads or coupons or subscribing to a newsletter.
- Should we use Unique Visitors (browsers) or Visits?
- Common for a unique visitor to visit the same page many times (purchase).
  - Can you think of an example?
- In this context conversion rate measures the process of converting a visitor into a buyer.
- Conversion rate can be calculated automatically by integrating analytics software with shopping carts.

# Engagement

- This is a qualitative metric hard to measure. The definition can be fuzzy.
- Most analytics software will track event and visit duration.
- This does not provide any information about the quality of engagement during that visit.
- What it matters is the time a visitor spends on a web site **with** engagement.
- The challenge is to distinguish between
  - Reading the information on a web page or
  - Looking for the information and not being able to find it.
- Example: Google Analytics tools allow us to research engagement
  - In-Page Analytics (visual assessment of how users interact with your web pages)
  - Behavior Flow analytics (visualizes the path users traveled from one page or event to the next).
- Both of these contribute to Engagement statistics.
- More on Google Analytics in the next session.

