eTail CRM Summit: Mining Customer Data

Ronny Kohavi, Ph.D.

Senior Director and Chief Evangelist of Business Intelligence

ronnyk@bluemartini.com

Elevating the Customer Experience™

Talk Outline

- The Web is an Experimental Laboratory
- Measurement and Collection of Data
- Analysis
- Action

This talk has many examples. They are all based on real data that I was personally involved in analyzing

The Oracle

- **How much would you pay for an Oracle who could tell you**
  - **Which of three creative designs is the best one to use in a campaign?**
  - **What your customers are looking for but not finding?**
  - **What is the next Pokemon (e.g., hot item)?**

The R&D Lab

- The web is an experimental laboratory
- As a channel, the web may generate only 5% of your revenues
- As a lab, it can help you
  - Test Campaigns
  - Test new product introductions
  - Identify products customers are searching and not finding
  - Identify cross-sells
- Many findings in the lab will carry over to other channels

The Web: Good

- **The web is a special channel with huge advantages for mining data**
  - **"Perfect" data collection is possible**
    - **Every interaction (page view, search, form) can be recorded**
    - **Views and transactions (e.g., purchases) are easily correlated**
    - **A lot of data is available quickly**

      Even a small site selling 5 items an hour will have 1.6 million page views after the first month
  - **Electronic collection more reliable other channels where data is entered by hand**
  - **Actionable – easy to change things online**

- **The web is also a place for customers to do research.**
  Forrester claims that "29% of the online population researched purchases online to buy them offline" and that the web "will influence 26% of total retail sales" by 2005

Web: the Bad and Ugly

- **Like every idea, there are limitations**
  - **The web is a biased sample of your customers**

    Not everyone is online (but more and more are)
  - **Online behavior is different**

    People will rarely buy an expensive suit online.

    However, large appliance sales is a category growing quickly (you pay for delivery of the dishwasher anyway, might as well research and get it cheaper online)
  - **A good web site is harder and more expensive to build than what people expect**
  - **A lot of automated spiders/robots crawl the web**

Talk Outline

- **The Web is an Experimental Laboratory**
- **Measurement and Collection of Data**
- **Analysis**
- **Action**

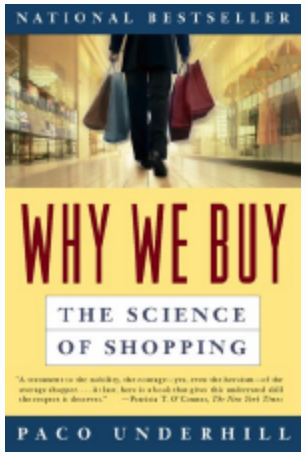
Measure and Record Data

- **Why?**
  - Data is the input to analysis
  - Without analyzing data, you are flying blind
- **How much?**
  - Ideally, everything that could be of value
  - In reality, it is an economic question.
    You trade off knowledge and insight versus collection cost
    - Transactions (e.g., purchases) are always recorded
    - Contacts with salespeople are sometimes recorded, sometimes in the salesperson's black book
    - Behavior (e.g., physical browsing) is rarely recorded

Recording Behavior

- **Excellent example of collection and analysis in physical stores**

  Why We Buy: the Science of Shopping by Paco Underhill



- **Human trackers fill logs**

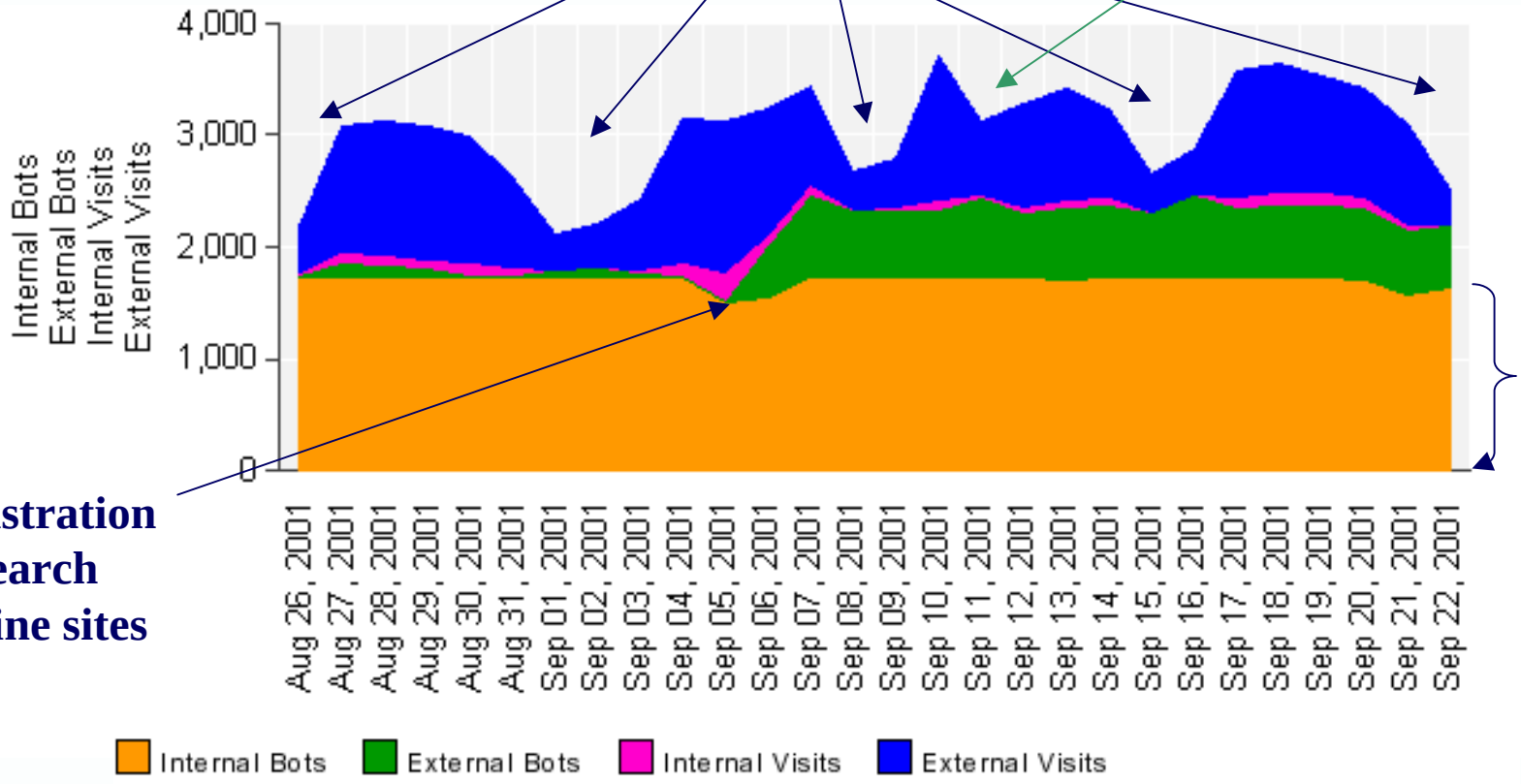
  She's in the bath section. She's touching towels.
  Mark this down -- she's petted one, two, three, four of them so far.
  She just checked the price tag on one. Mark that down, too.
  Careful, her head's coming up -- blend into the aisle.
  She's picking up two towels from the tabletop display and is leaving the section with them.

- **EnviroSell Inc. goes through 14,000 hours of store videotapes a year to do behavioral research**
- **The Web automates much of this and makes it economical for YOU to see similar data**

Example: Web Traffic

Sept-11
Note significant drop in human traffic, not bot traffic

Weekends

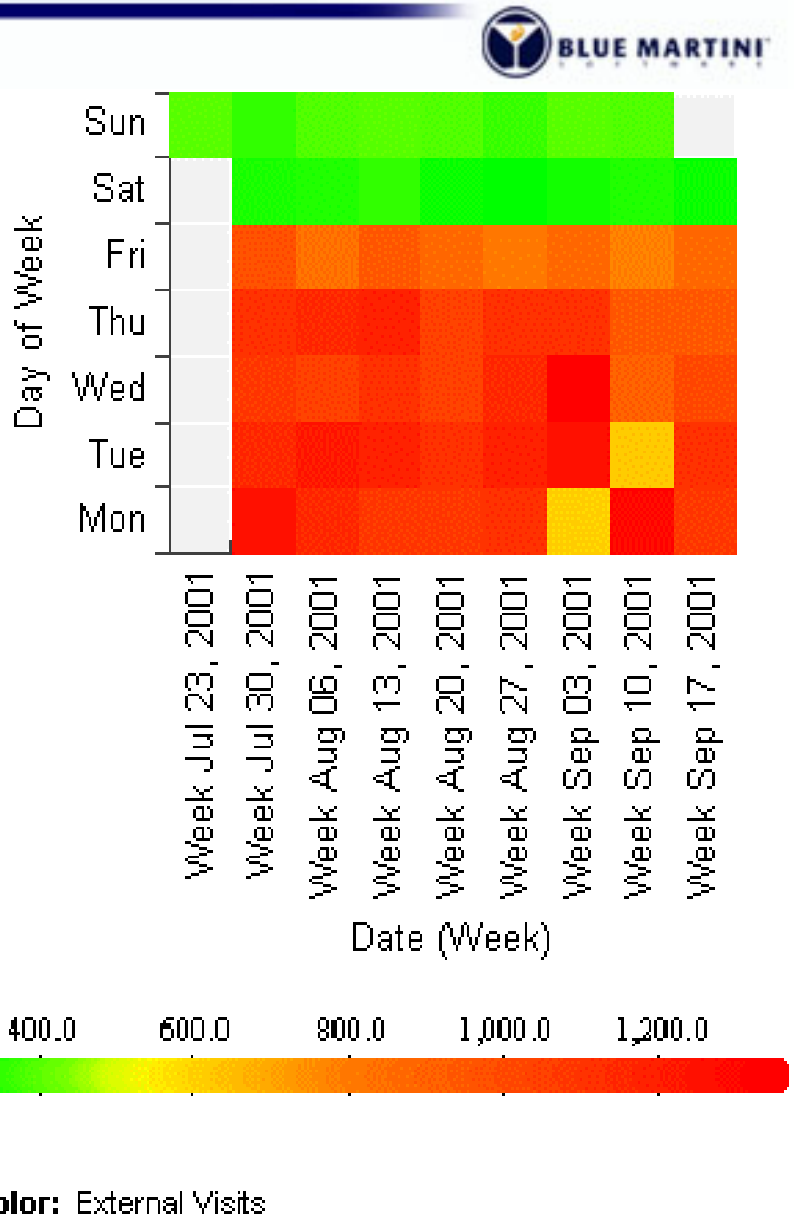


Internal Perfor-mance bot

Registration at Search Engine sites

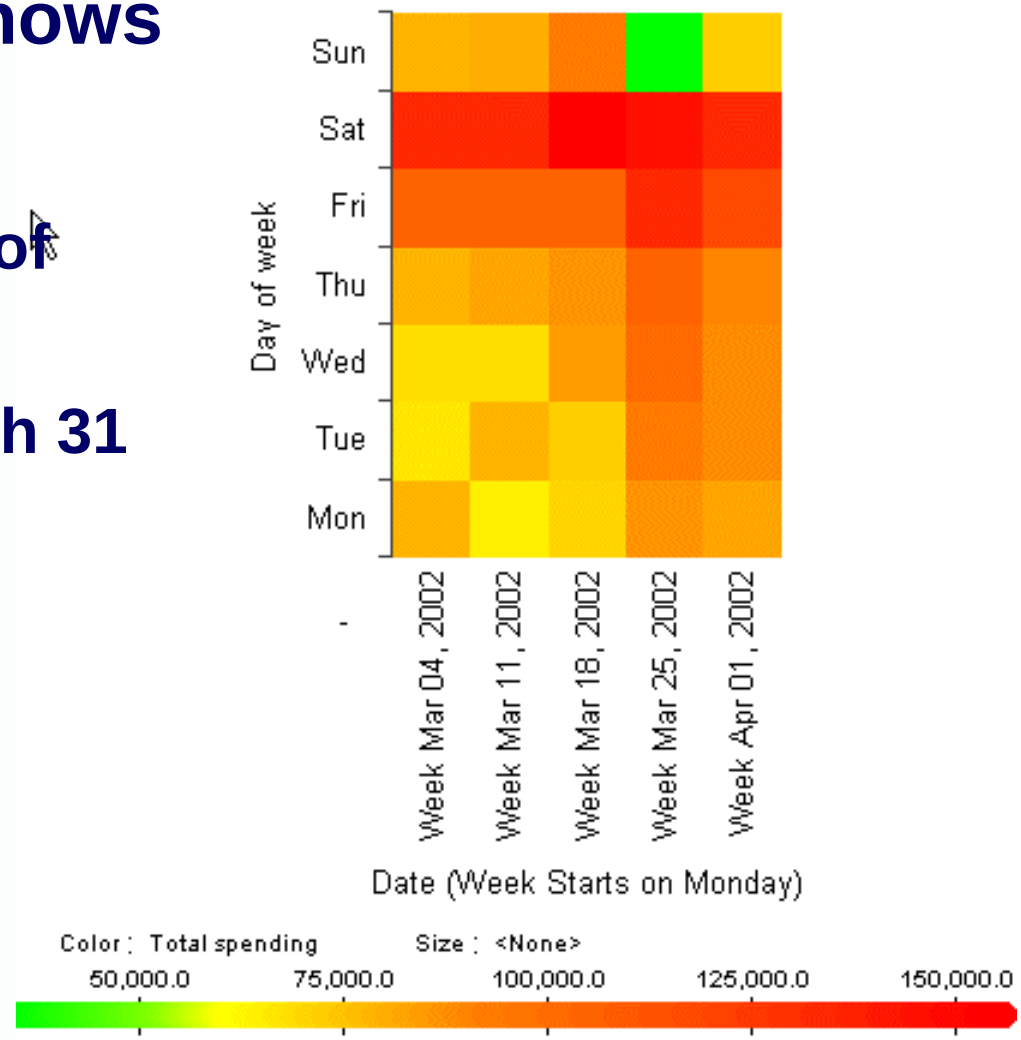
Heat Maps for Day-of-Week

- Visualization can help see other patterns in the same data
- Use color to show traffic
  - Green is low traffic
  - Yellow is medium traffic
  - Red is high traffic
- Observations
  - Weekends are slow
  - Patterns
    - Sept 3 Labor day in yellow
    - Sept 11 in yellow
    - Reduced traffic after Sept 11
    - Reduced traffic Fridays



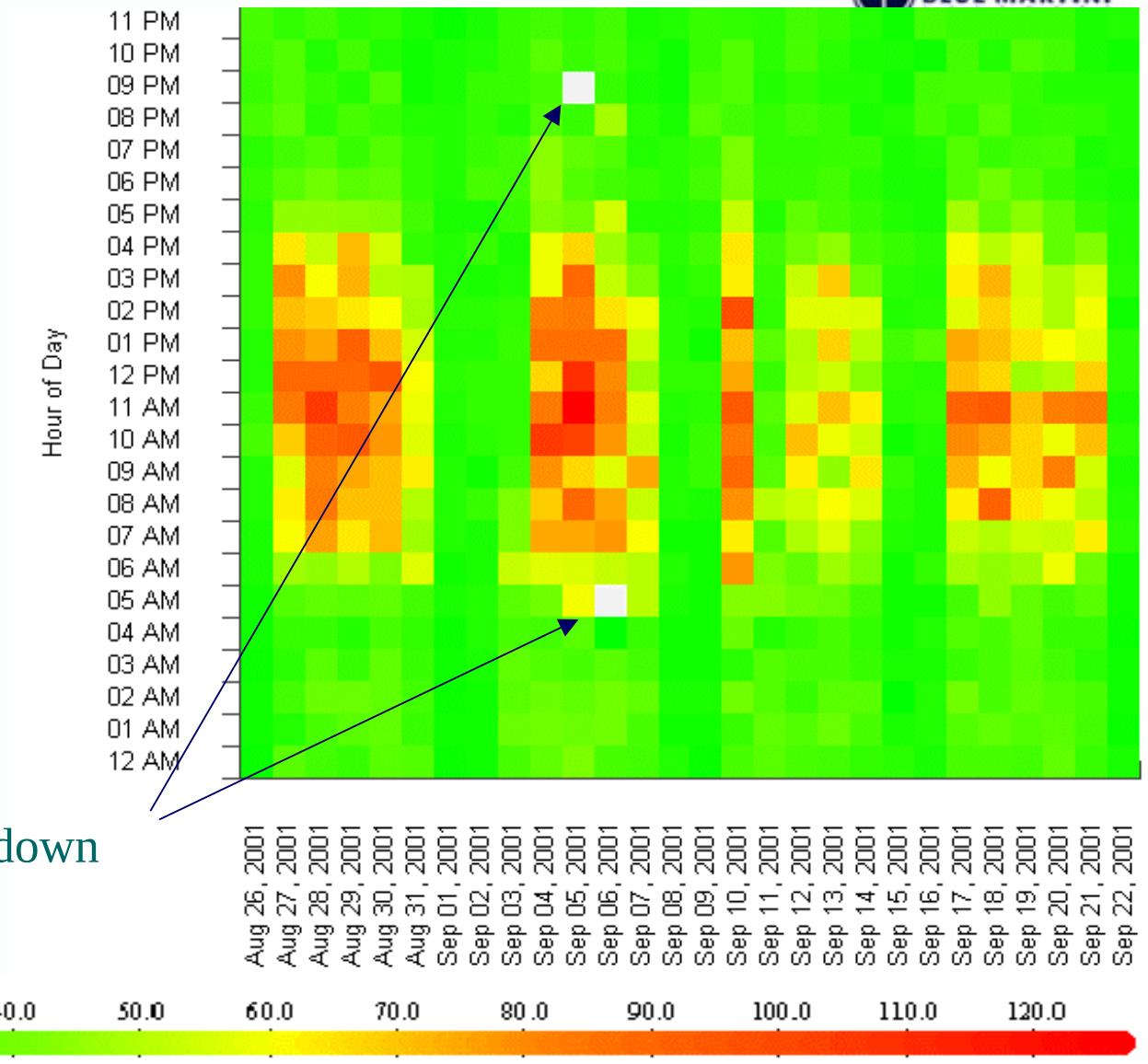
POS Data Example

- **Sales data from POS shows**
  - **Strong Friday/Saturday**
  - **Heavy purchases week of March 25th**
  - **Very low revenues March 31**



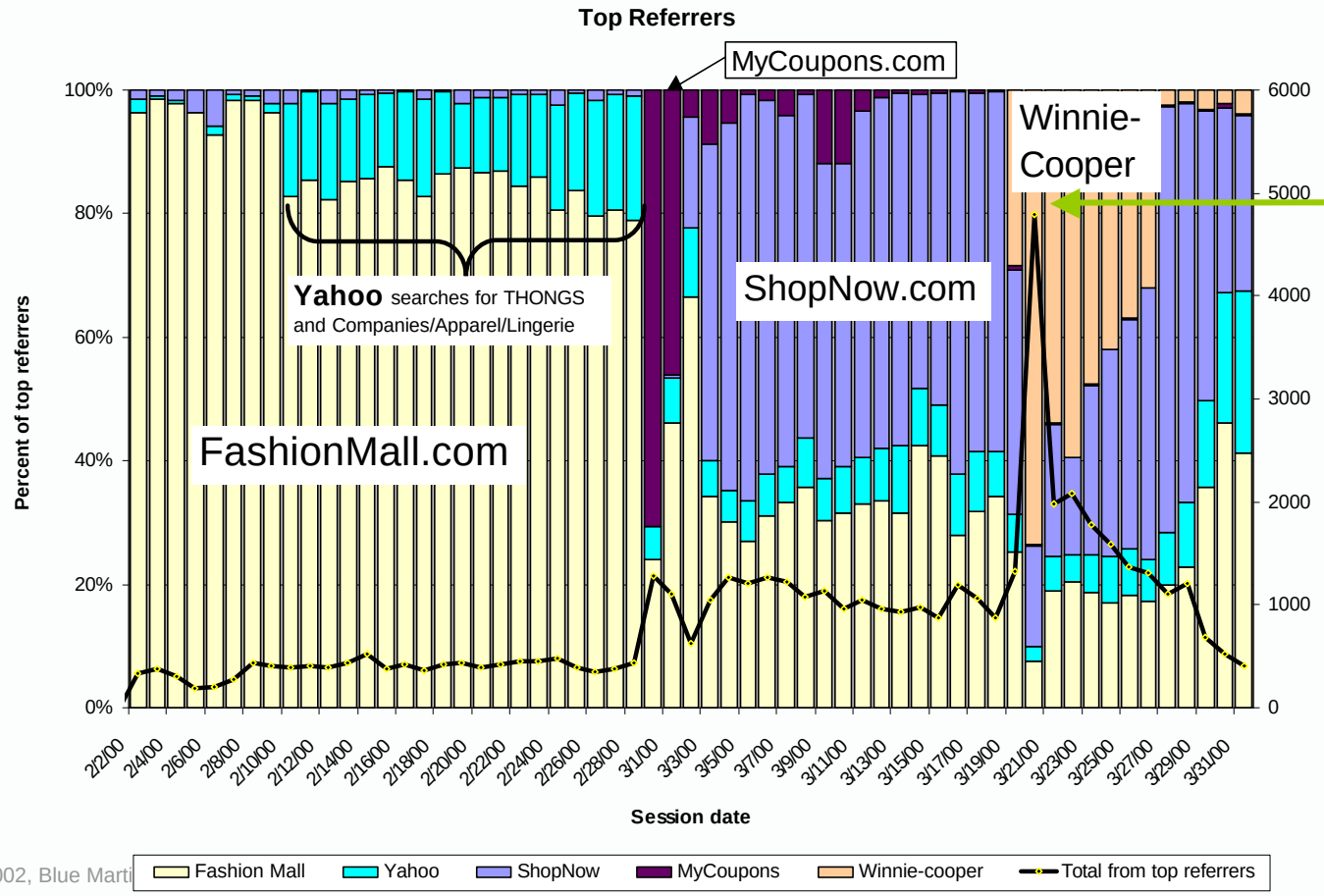
Drill-Down to Hour

- Helps determine best time for maintenance
- Note Sept 11 effect and its effect for rest of week



Teaser - Who is Winnie?

Referring site traffic for a leg-wear and leg-care web retailer.
Who is Winnie Cooper? What can you do about it?



Note spike in traffic

Answer to Teaser

- Winnie-cooper is a 31 year old guy who wears pantyhose
- He has a pantyhose site
- His website averages 15,000 - 20,000 visitors a day
- Large number of visitors came from his site
- **Actions:**

- Make him a celebrity and interview him about how hard it is for a men to buy pantyhose in stores
- Personalize for XL sizes

Visitors from Google

- **Search engines help you identify what keywords people use to find your site**
- **Search on Google that came to Blue Martini Software**
  - **B2C website**
  - **case study retailing apparel**
  - **ERP case study**
  - **retail CRM**
  - **most profitable retail websites**
  - **CRM Retail Software**
  - **consumer loyalty**
  - **grocery software**
  - **campaign management**
  - **ECRM**
  - **data mining application case study**

ale/Clickthroughs

- **The number of sessions is a simple metric.**
  More interesting is to correlate sessions with purchases and behaviors
- **Client 1 search referrals**
  - Google:5% of traffic, $0.61/clickthrough
  - Yahoo: 3% of traffic, $0.66/clickthrough
  - MSN : 2% of traffic, $0.51/clickthrough
  - AOL: 1% of traffic, $0.68/clickthrough
- **Client 2 search referrals**
  - MSN: 2% of traffic, $1.98/clickthrough
  - Yahoo: 2% of traffic, $1.89/clickthrough
  - AOL: 1% of traffic, $3.54/clickthrough
  - Google:1% of traffic, $2.52/clickthrough
- **Clear ROI. How much are you paying per clickthrough?**

What to Collect on the Web

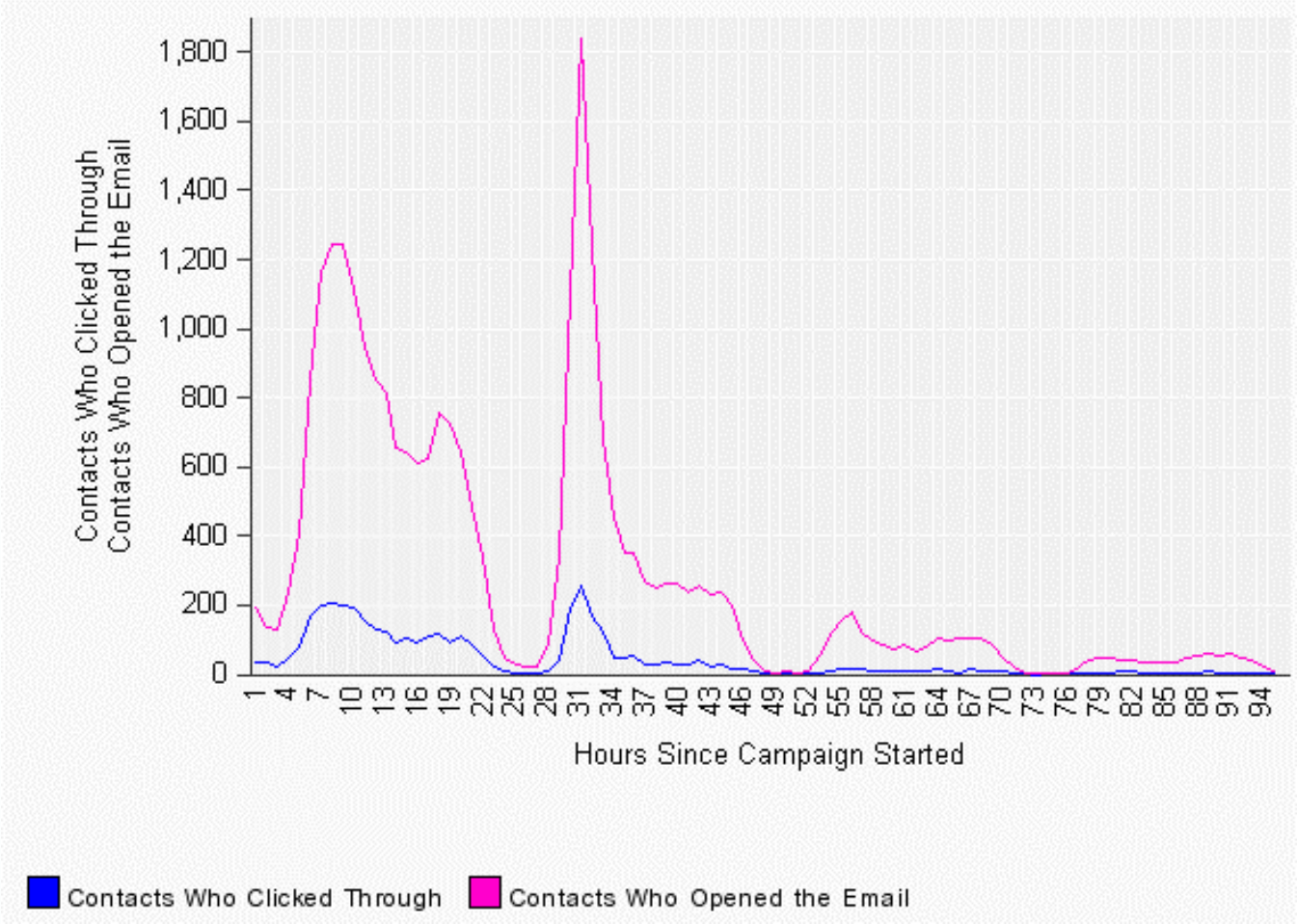
- **Some things to collect on the web that are non-standard**
  - **User local time zone**
    Understand when users are browsing in THEIR time zone
  - **Screen resolution**
    An excellent surrogate for techies (high res).
    Appears as a an important factor in analyses
  - **Events (add to cart, remove from cart, registrations, searches)**
  - **Errors on forms**
    If many make mistakes, fix the question
  - **Spider/bot tricks (hidden link, zip support)**

E-mail Campaigns

- **When doing an e-mail campaign, make sure to**
  - **Personalize the e-mail**
  - **Make every link unique (redirection link) to allow for user identification on clickthrough**
  - **Track revenues by clickthroughs and visits by recipients to site following a campaign**
  - **Use unique coupon codes to identify user across all channels (e.g., use of e-coupon at stores)**
  - **Put a “web bug” (single-dot image) that is retrieved from the server on e-mail opening**
    Allows computing the e-mail “open rate”

E-mail campaign opening/clickthroughs

- **E-mail campaigns are over in 3-4 days**
- **Spikes correspond to mornings**



Campaign sent at 3AM EST

Talk Outline

- The Web is an Experimental Laboratory
- Measurement and Collection of Data
- Analysis
- Action

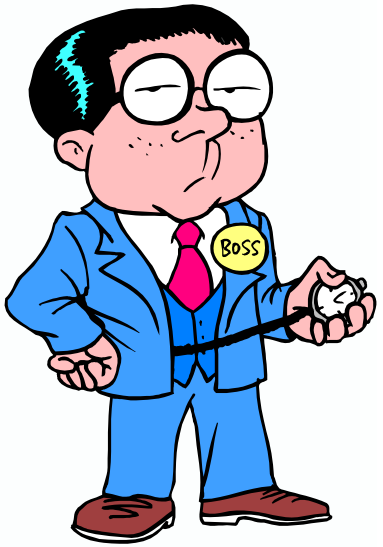
Separate Analysis System

- **Analysis should be performed on a separate copy, not on the operational system**
  - **Do not kill the performance of your operational system**
  - **Data structures (e.g., database schema) are different**
    - **Operational side is designed for small queries/updates**
    - **Analysis requires massive queries**

    Build the appropriate schema (e.g., star schema) for your data warehouse
  - **Work with stable data that does not continually change. Use alerts to trigger immediate action for basic metrics that are out of range at the operational side**

Data Transformations

- **80% of the time spent in data analysis is typically spent transforming data**
  - **Expect and plan for pain and effort**
  - **Use the right ETL tools**
  - **Automate transfers**
  - **Look for software that has automated transformations for your needs, reducing the 80% dramatically (e.g., Blue Martini provides automatic transformations from the e-commerce site to the Decision Support System)**

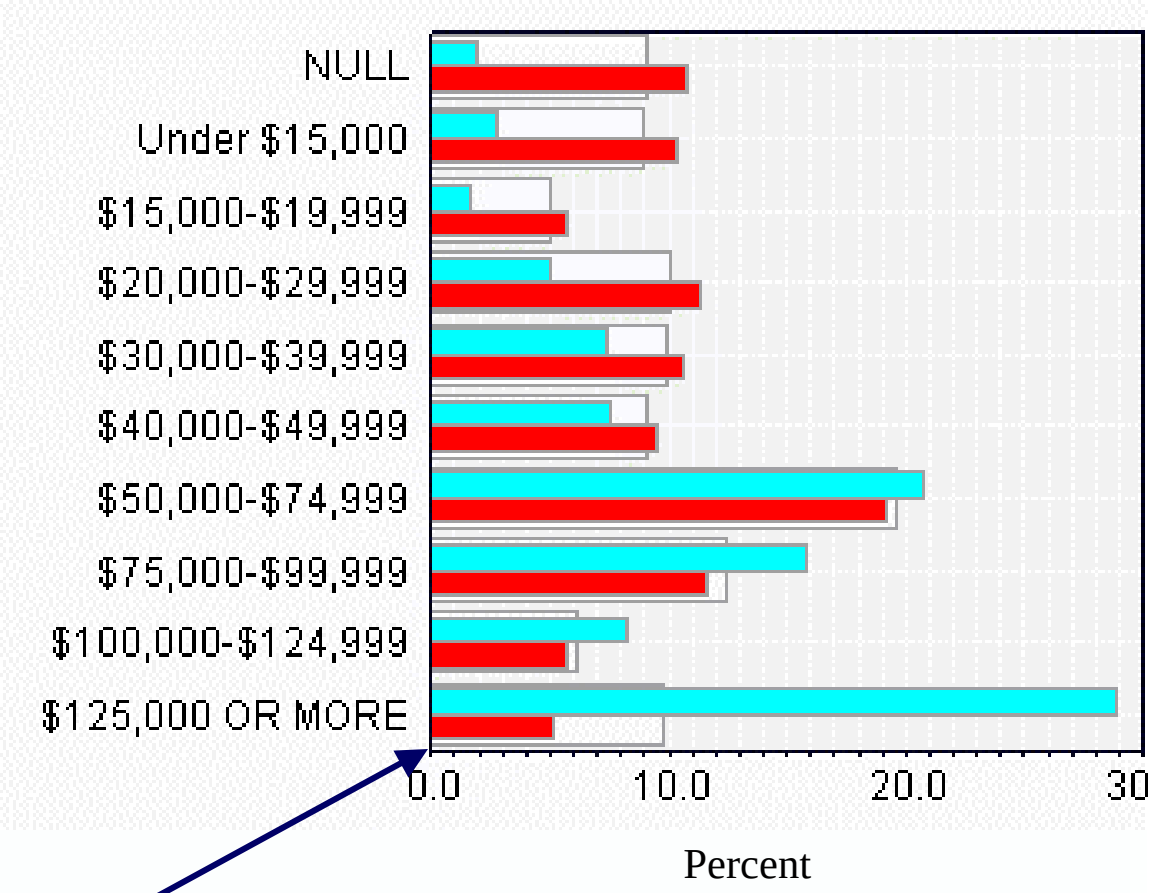


Demographic Overlays

- **Combine information about your customers from all possible sources**
- **A great source that is often overlooked is demographic overlays**
- **Very easy and cheap (8 cents/name) to get basic attributes like gender, income, profession, own/rent, car type, etc.**
  Note: not very reliable per person, but good for averages and segmentation

Example - Income

- Graph showing incomes for a company that targets high-end customers based on POS purchases
- Income of their customers in blue
- The US population in red

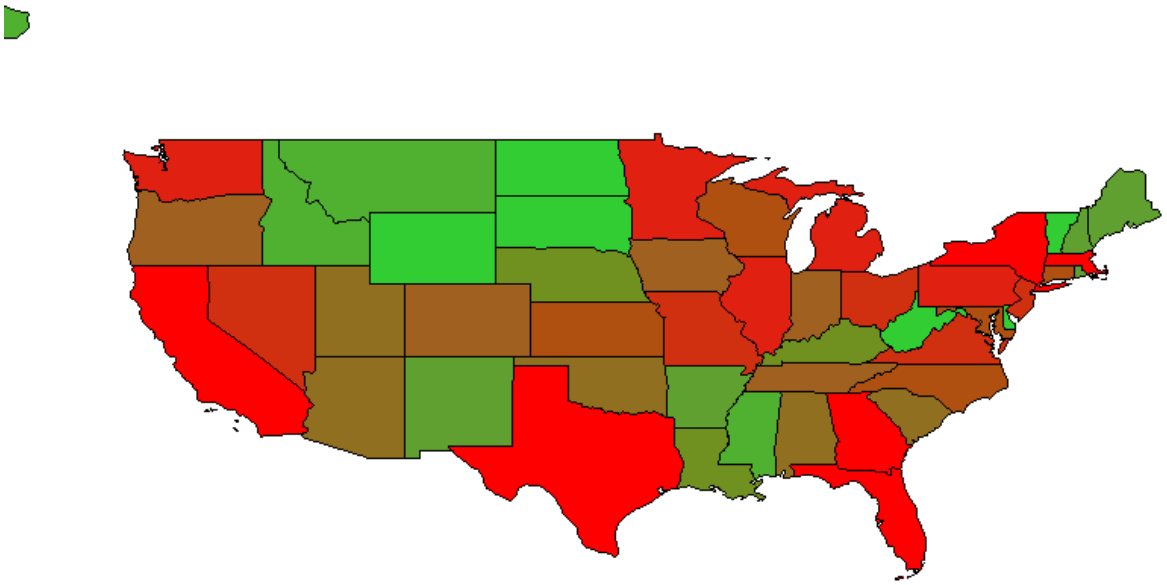


Note highest bracket

Look at Geography

BLUE MARTINI

Customers by state



USA

5,545 to 12,671	(6)
3,293 to 5,545	(5)
1,728 to 3,293	(5)
1,171 to 1,728	(5)
884 to 1,171	(5)
601 to 884	(5)
481 to 601	(3)
266 to 481	(6)
202 to 266	(5)
106 to 202	(6)

E-Metrics Study

- Project to understand behavior on the Web
- Based on data collected during the 2000 holiday season from multiple Blue Martini clients
  - US and European sites
  - B2B and B2C sites
  - Several different industry verticals
- Results based on
  - More than 1,000,000 online visits
  - More than 500,000 online visitors
  - More than 50,000 registered customers
  - Acxiom random sample of 20,000 people from US

Spiders, Crawlers, and Robots

- Spiders/crawlers must be filtered to avoid skewing statistics

- Main types:
  - Search engines
  - Content grabbers (email scanners)
  - Browser crawlers (IE site caching)
  - Site monitors (Keynote, Patrol)
- What percentage of visits are robots?

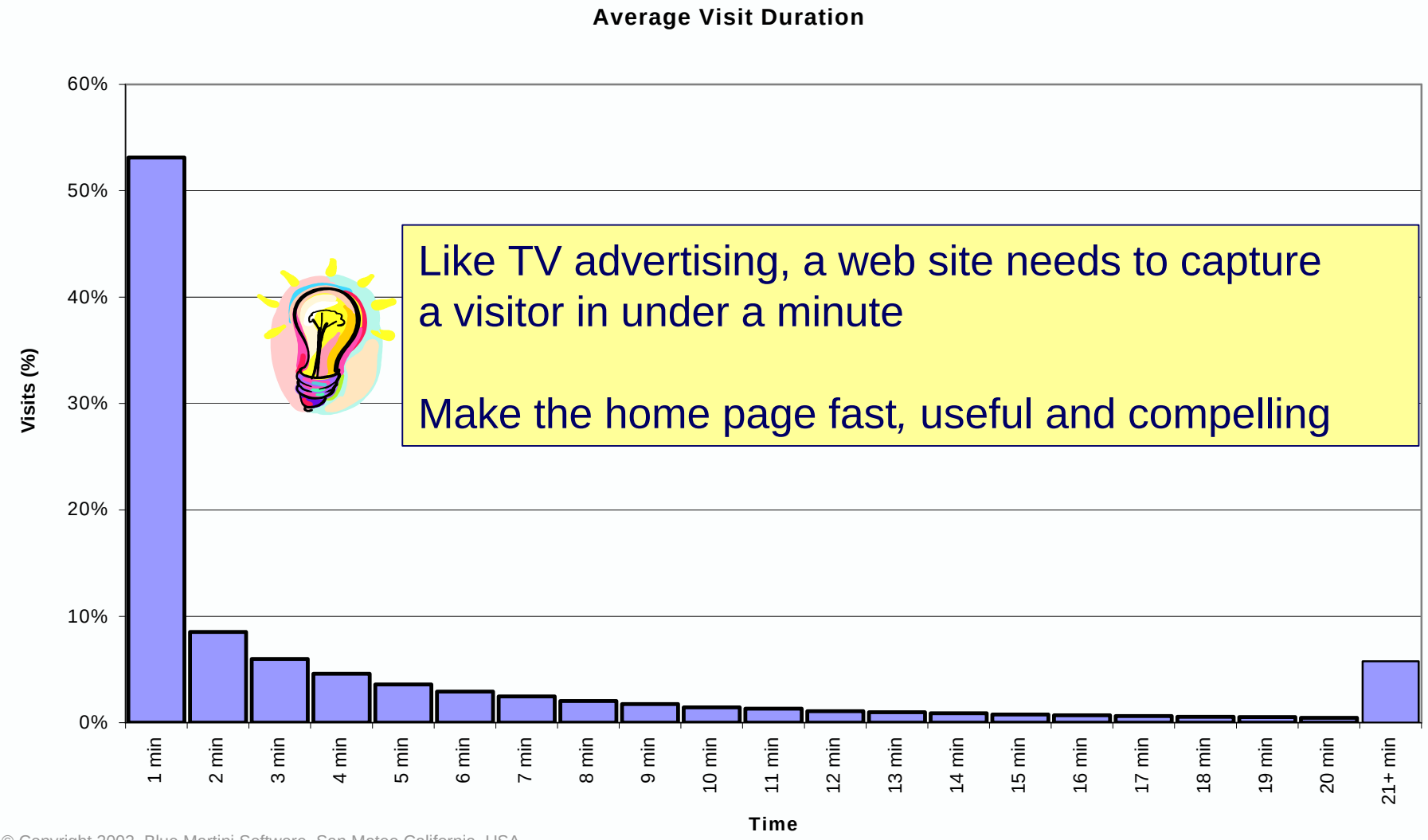
About 25%

Visit Statistics

- An average visitor
  - Views 10 pages
  - Spends 5 minutes on the site
  - Spends 35 seconds between pages
- An average *purchasing* visitor
  - Views 50 pages
  - Spends 30 minutes on the site
- Very consistent across multiple web sites

Visit Duration

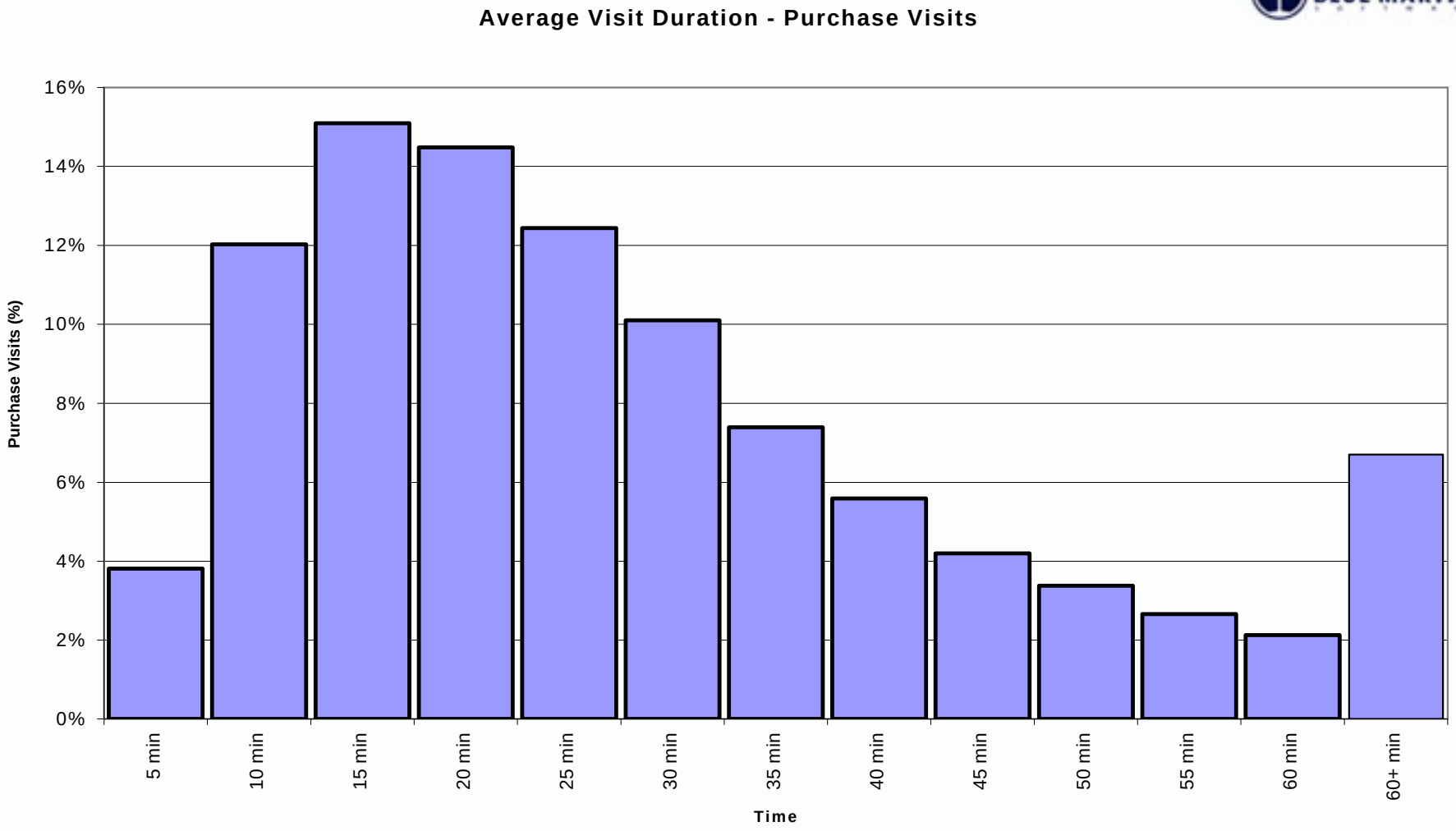
50% of visitors leave in under a minute



Like TV advertising, a web site needs to capture a visitor in under a minute

Make the home page fast, useful and compelling

Purchasing Visit Duration



Note scale in 5-minute increments

Privacy

- 92% of Americans are concerned (67% very concerned) about the misuse of their personal information on the Internet.

  - FTC Report, May 2000

- 86% of executives don't know how many customers view their privacy policies.

  - Forrester Report, November 2000

- Q: What percentage of visitors read the privacy statement?
- **A: Less than 0.3%**

Consumer Demographics

- **Using Acxiom, we compared online shoppers to a sample of the US population**
  - **People who have a Travel and Entertainment credit card are 48% more likely to be online shoppers (27% for people with premium credit card)**
  - **People whose home was built after 1990 are 45% more likely to be online shoppers**
  - **Households with income over $100K are 31% more likely to be online shoppers**
  - **People under the age of 45 are 17% more likely to be online shoppers**

Search

- Search correlates with better customers
- For a large site,
  - A visit with search is worth 54% more than a visit without search
  - Successful searches are key
    - If the last search failed, the conversion rate was 3.48%
    - If the last search was successful, the rate was 6.22%

Example - Search Keywords

- **On one of Blue Martini's client sites (sports-related), the top searched keywords were:**
  - **Baseball**
  - **Video**
  - **Softball**
  - **Volleyball**
  - **Pins**
  - **Equestrian**
  - **Videos**
  - **Posters**
  - **Music**
  - **Poster**

What is common to the words in red?

Example - Search Keywords

- **On one of Blue Martini's client sites (sports-related), the top searched keywords were:**
  - Baseball
  - Video
  - Softball
  - Volleyball
  - Pins
  - Equestrian
  - Videos
  - Posters
  - Music
  - Poster

Actions for failed searches:

- **Define synonyms in the search thesaurus**
- **Support misspelled words**
- **Expand merchandise assortments based on failed searches**

Across multiple sites, about 10% of searches fail

Failed Searches - Example

- **Consumer Reports magazine tests consumer products**
- **Here are the top failed searches**

karaoke (1.39% of failed searches)
atv (1.12%) - need synonym (All Terrain Vehicle
guitar, guitars (0.81%)
abtronic (0.49%)
boombox (0.49%)
cdrw (0.48%)
snowthrower (0.46%)
webcam (0.39%)

Total = 5.6% of failed searches

- **Action: failed searches provide excellent guidance to management about products that are interesting to consumers but not yet covered by the magazine**

Modeling

- **Use prediction models (e.g., classification) with two goals:**
  - **Comprehension. Some models, such as rules and trees are easy for business users to understand and lead to insight**
  - **Scoring. Assign scores to customers based on their propensity to buy something or behave a certain way (e.g., heavy spender). Use scores for personalization**
- **Use market basket analysis (associations) to suggest cross-sells**

Talk Outline

- The Web is an Experimental Laboratory
- Measurement and Collection of Data
- Analysis
- Action

Test, Test, Test

- Act often and test the effect
- Decide on automated actions for events, such as
  - Purchases,
  - Lifestyle changes (e.g., wedding),
  - Household moves, and
  - Service requests

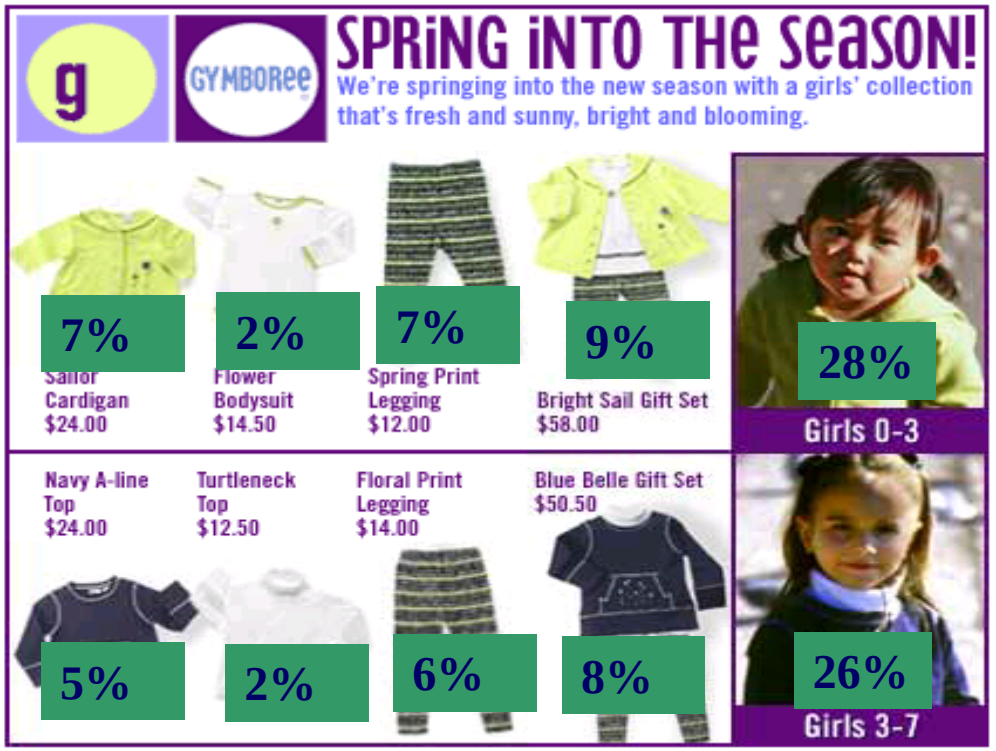
  More on this in later talk by Monte Zweben
- Test different campaigns on a sample before deciding the one to use

Example: Campaign Mailer

- Gymboree sent seven different e-mails
- Track which e-mails more effective
- Track where people click
- Lifestyle images
  - Two better than one

- Use darker colors

Example for illustration only and does not show actual percentages



Harley-Davidson

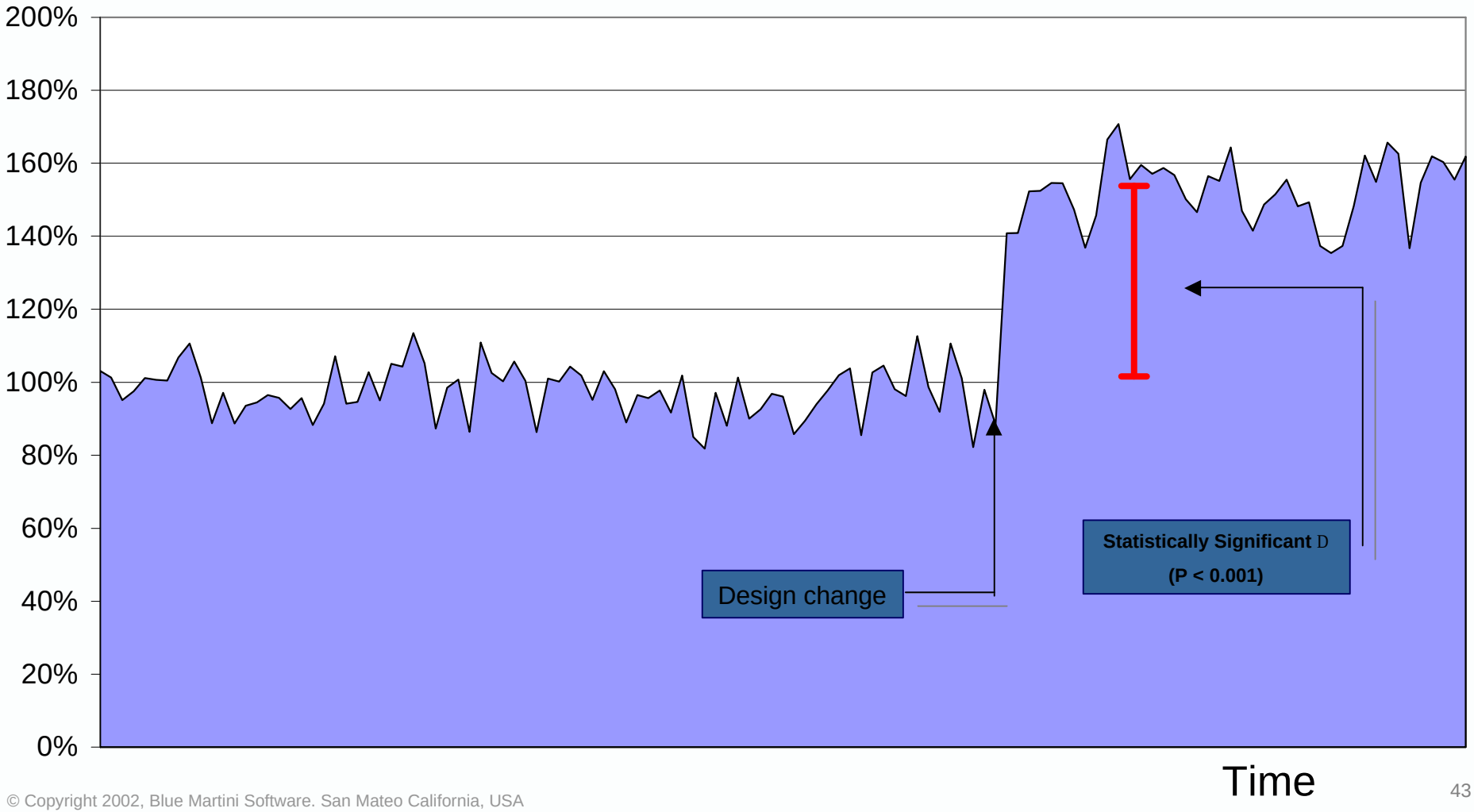
- **Harley Davidson has very loyal customers.**

  (So loyal they tattoo the corporate brand name and logo on their body.)

- **However, certain processes on site were too complicated even for these loyal customers**
- **Blue Martini Analytic Services analyzed their site and made recommendations for improvements**

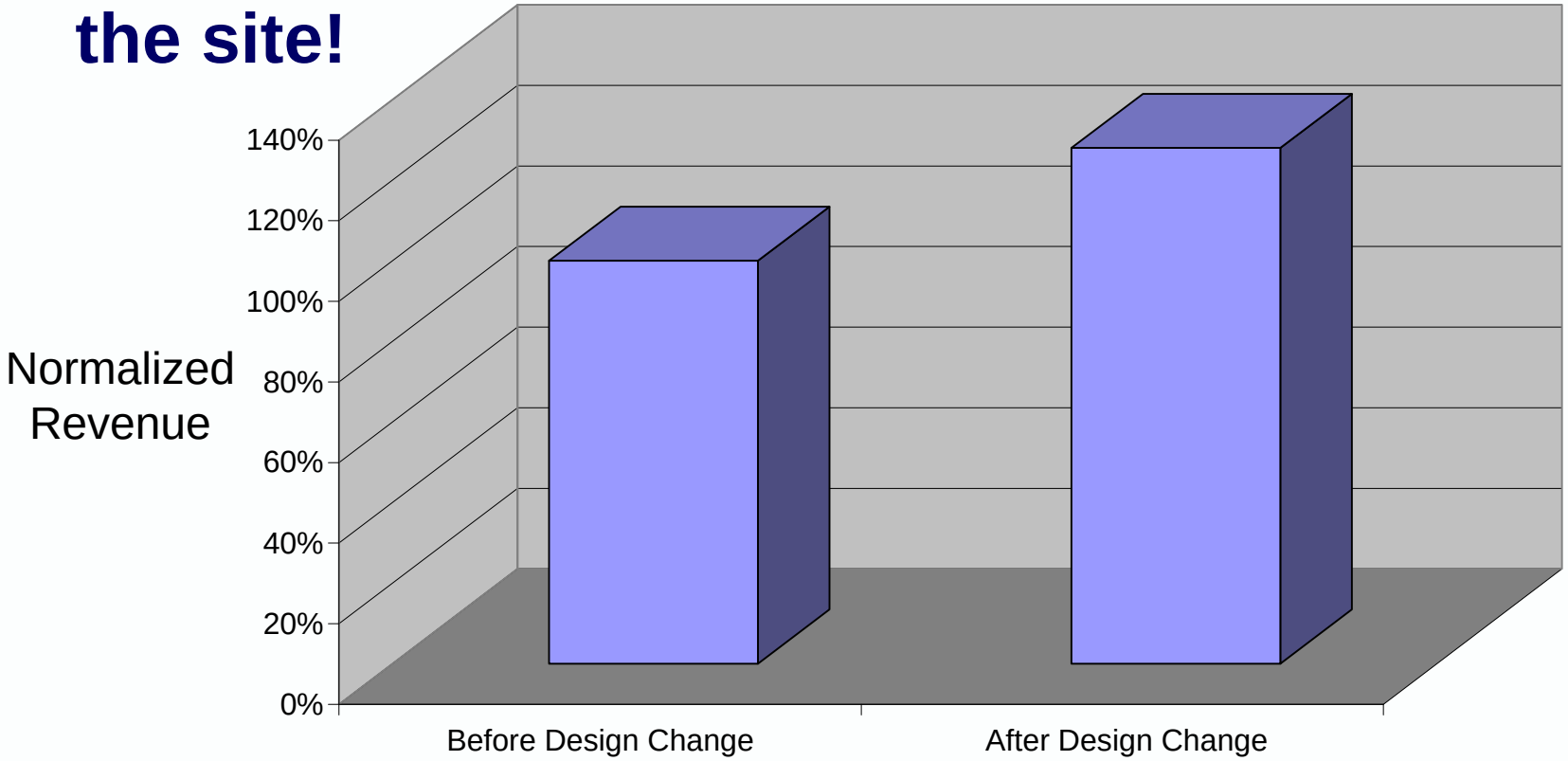
Significant Improvements After Action

Over 50% increase in sessions initiating and completing process



Financial Impact

- **Increase in revenue of over 120% for process users**
- **Nearly 30% increase in overall revenue from the site!**



Summary

- **The Web is an Experimental Laboratory**
  - **The web is a unique channel with perfect data collection (an e-metrics study for physical stores is much harder)**
  - **Use the web to analyze behavior, detect trends, test ideas, then apply at other channels**
  - **Try a lot of stuff and keep what works**
- **Measure and collect more**
  - **Definitely cost effective on the web**
  - **Attempt to get more at other channels and integrate activities from all channels for a panoramic view**
- **Analyze**
  - **Find gold in your mountain of data – mine it!**
  - **Use visualizations because they are easier to understand**
- **Action**
  - **Insight leads to action**
  - **Make continuous changes and measure their effect**

To find knowledge nuggets in your data, contact Blue Martini Software

We provide software and/or analytic services

Ronny Kohavi, ronnyk@bluemartini.com

We wish to thank IBM for co-sponsoring this event

Elevating the Customer Experience