

eMetrics Study

Llew Mason, Zijian Zheng, Ron Kohavi, Brian Frasca

Blue Martini Software

`{lmason, zijian, ronnyk, brianf}@bluemartini.com`

December 5th 2001

© 2001 Blue Martini Software



1. Introduction

Managers of e-commerce sites have strategized to build dynamic, personalized, content-rich sites in an effort to attract new customers, increase wallet share, and streamline navigation. The need to manage the success of these initiatives has resulted in an emergence of new measurements or, “e-commerce Metrics”. Managers are using these metrics to determine the value of their e-commerce offering, to gauge the return on their development investments, and to monitor system diagnostics.

In early 2001, Blue Martini Software approached several clients asking them to be part of an extensive study to generate a set of eMetrics using their transactional, customer and clickstream data. There were two main goals of this study:

1. To develop new e-commerce metrics that can be compared across different sites irrespective of the business model, industry vertical or geographic location.
2. To provide Blue Martini clients with information on how the browsing and buying behavior of their customers differs from the behavior of customers on other sites.

We obtained data collected over the 2000 holiday season from five clients’ e-commerce web sites running Blue Martini’s e-commerce software. This report is the result of that study.

The goal of generating *comparable* eMetrics was made more difficult by the diversity of clients involved in the study. The five clients’ site designs, geographic locations and business models varied widely, including B2C and B2B business models, several different industry verticals, and sites based in both the US and Europe. Due to these differences, eMetrics like the total sales or the total number of visits or visitors don’t make sense when compared across different clients. For this reason, this study includes eMetrics based on quantities like conversion rates, think time, visit duration, and shopping cart and search behavior.

We also generated metrics comparing the demographics of the customers in the study to a random sample of the US population. The customer data obtained from each site in the study was enhanced using syndicated data provided through the Acxiom Data Network. This introduced new customer attributes such as Gender, Occupation, Age, Marital Status, Estimated

Income, and Home Market Value. These attributes were then aggregated over all of the customers from each site. This provided comparative site-wide demographics without maintaining any information about individual customers' attributes.

As mentioned above, all analysis was restricted to the 2000 holiday season between 1 November 2000 and 31 December 2000.

The remainder of this report is organized as follows. In Section 2 we describe the data processing performed in order to generate the eMetrics. Then, in Sections 3 and 4, we give the eMetrics results and comparative demographics averaged over all of the clients that were involved in the study. Finally, in Section 5, we give detailed definitions for each of the computed eMetrics.

2. Data Processing

Each client provided us with backups of their databases and some technical information about their Blue Martini implementation and configuration. For each client we then:

1. Restored the provided databases (MAIN, STORE and CLICKSTREAM).
2. Upgraded the databases to v4.0p1 of Blue Martini.
3. Ran DSSGen to build the DSS data warehouse, pulling demographics from the Acxiom Data Network for the US based sites in the study.
4. Identified robots and crawlers and marked those sessions.
5. Cleaned up any known or discovered data quality issues.
6. Built data mining investigations using the Blue Martini Data Mining Manager to calculate the eMetrics and generate the comparative demographics.
7. Exported the results of the data mining investigations into Microsoft Excel in order to generate comparisons for each eMetric between each client and the average over all clients.

Robot/Crawler Detection

Robots or crawlers are software agents that visit sites on the Internet for various purposes. Most of these robots download a page and then follow the links on these pages. Robots are most often used for search engine catalog generation, site diagnostics and offline browsing. In our experience, robots account for a significant proportion of a site's traffic, but they are not customers or visitors whose behavior we want to analyze. To remove the influence that these sessions have over the calculated eMetrics we tagged sessions generated by robots according to the following criteria:

1. User agents whose name contains any of the following keywords 'Keynote', 'Patrol', 'WebTrends', 'Teleport Pro', 'Lycos', 'Google', 'BackRub', 'Slurp', 'Gulliver', 'Architext', 'Scooter', 'crawl', 'spider', 'bot' (but not 'robotic'), 'archive', 'Mozilla/3.01 (compatible;)', 'Mozilla/4.0 (compatible; MSIE5.0; Win32)', and 'Mozilla/3.01 [en] (Win95; I)'. The last three represent known robots disguised to look like valid user agents.
2. User agents with all visits consisting of only a single page view or more than 100 page views (must have more than 20 visits).
3. User agents with no referrer set for all page views (must have more than 5 visits and every visit must have more than a single page view).
4. User agents which don't have the word 'Mozilla' in the name and have never logged in (must have more than 200 sessions).

5. User agents which visited the ‘crawlers.jsp’ page (if available).

The generated list of user agents was then scanned manually to ensure that no valid browsers were included by the above criteria.

The eMetrics

The table below lists all of the 30 eMetrics computed in this study.

Sales by week
Visits by week
Visitor to registered visitor conversion rate
Visitor to registered visitor conversion rate by week
Registered visitor to buyer conversion rate
Registered visitor to buyer conversion rate by week
Visitor to buyer conversion rate
Visit to buy conversion rate
Visit to buy conversion rate by week
Visitor to repeat visitor conversion rate
Average visits before registration
Shopping cart abandonment rate
Percentage of visits using search
Percentage of failed searches
Percentage of visitors reading privacy statement
Average page views per visit
Visit duration (pages)
Average page views per purchase visit
Visit duration (pages) – purchase visits
Average time per visit
Visit duration (time)
Average time per purchase visit
Visit duration (time) – purchase visits
Average think time
Think time distribution
Normalized spending distribution
Average home page processing time
Average processing time

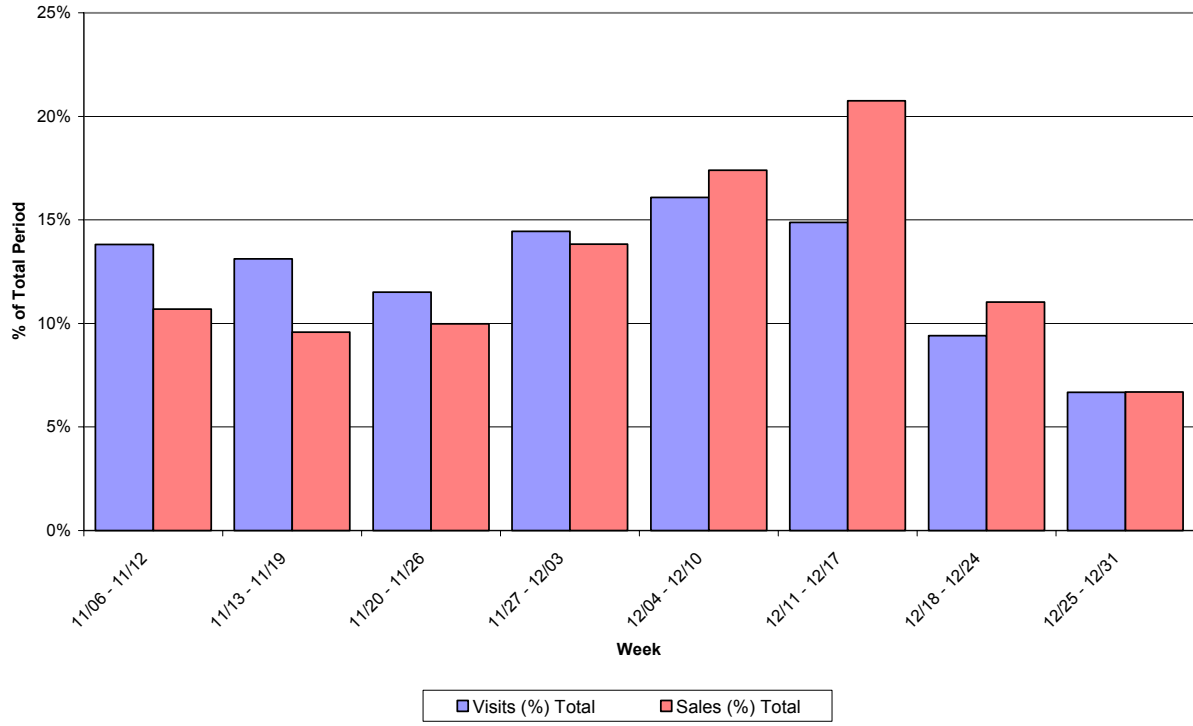
Percentage of visits due to robots/crawlers
Browser distribution

3. eMetrics Results

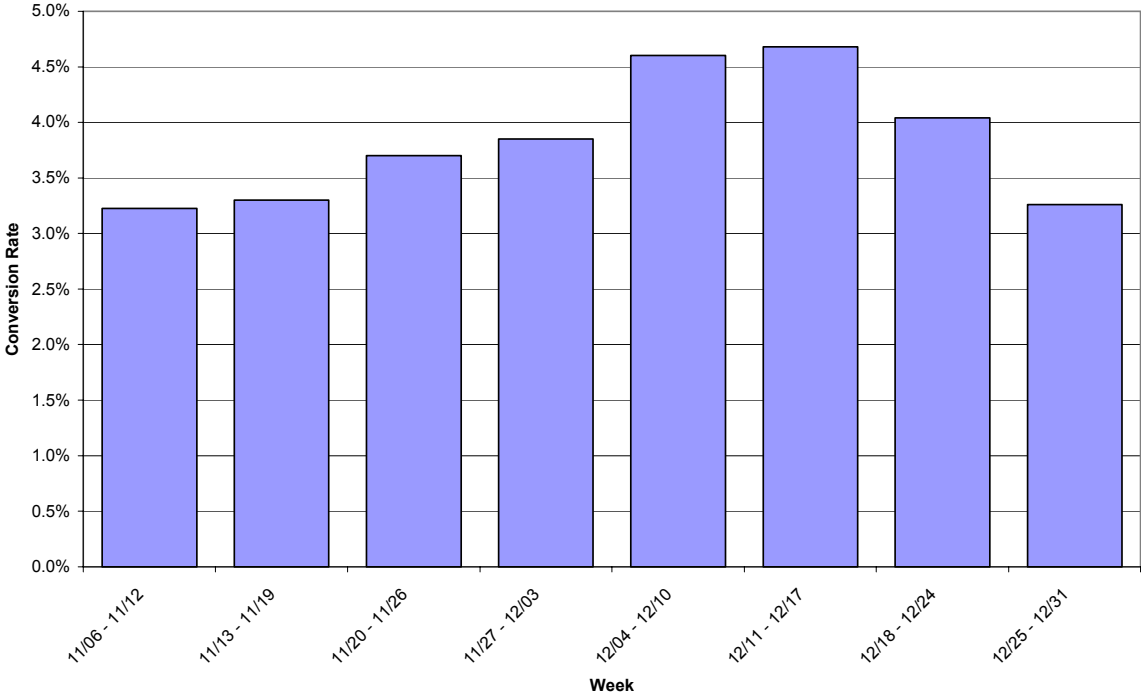
This section describes the results for each of the individual eMetrics averaged over all of the participating sites. Definitions of the eMetrics presented in this section are listed in Section 4.

	Average
Visitor to registered visitor conversion rate	5.0%
Registered visitor to buyer conversion rate	33.1%
Visitor to buyer conversion rate	1.7%
Visit to buy conversion rate	1.8%
Visitor to repeat visitor conversion rate	7.3%
Average visits before registration	5.9
Shopping cart abandonment rate	34.3%
Percentage of visits using search	6.2%
Percentage of failed searches	11.4%
Percentage of visitors reading privacy statement	0.2%
Average page views per visit	9.5
Average page views per purchase visit	50.7
Average time per visit	4.9m
Average time per purchase visit	28.0m
Average think time	34.9s
Average home page processing time	0.5s
Average processing time	1.2s
Percentage Bots	29.7%

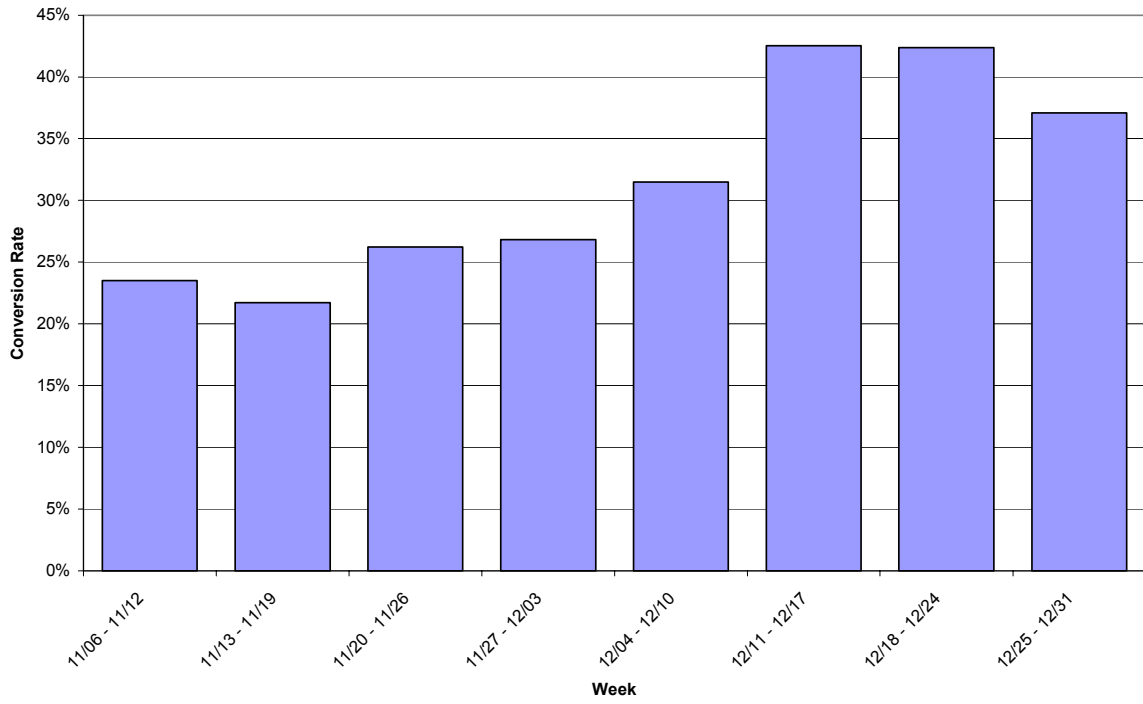
Average Visits and Sales by Week



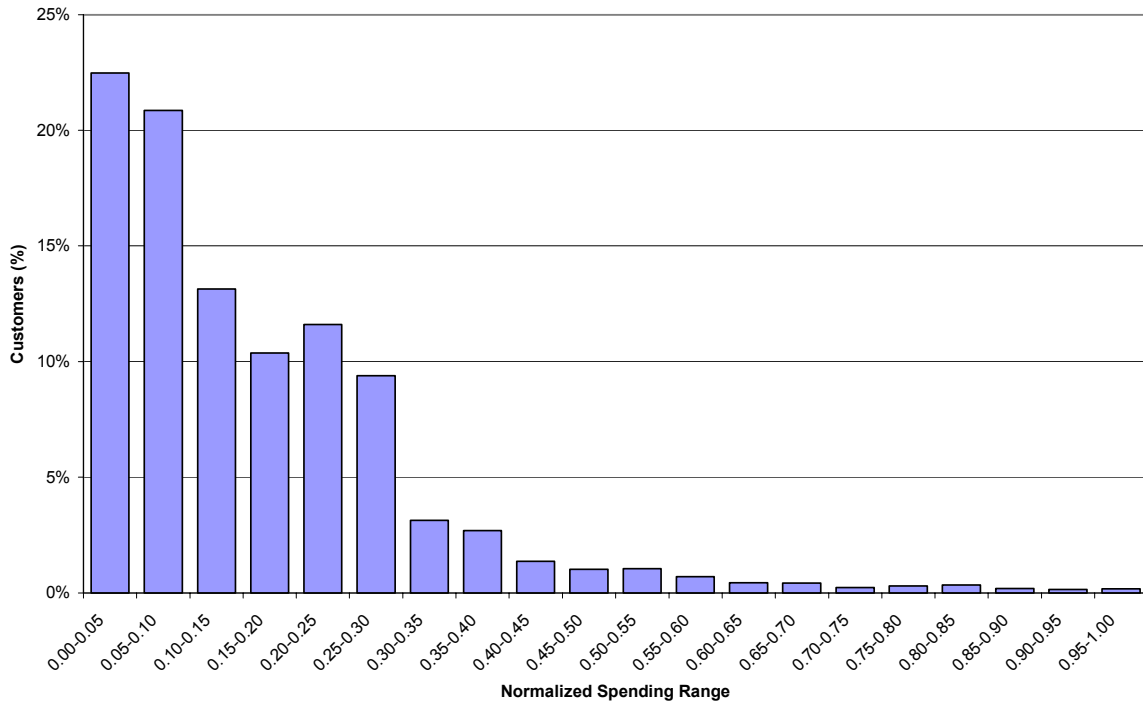
Average Visitor to Registered Visitor Conversion Rate



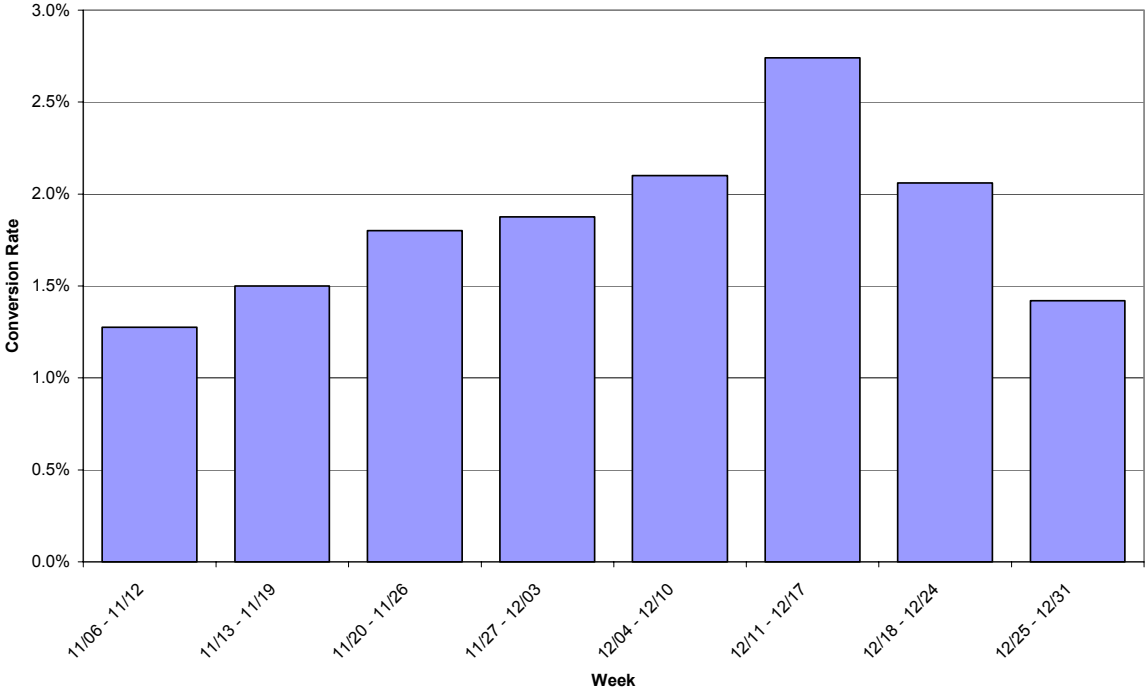
Average Registered Visitor to Buyer Conversion Rate



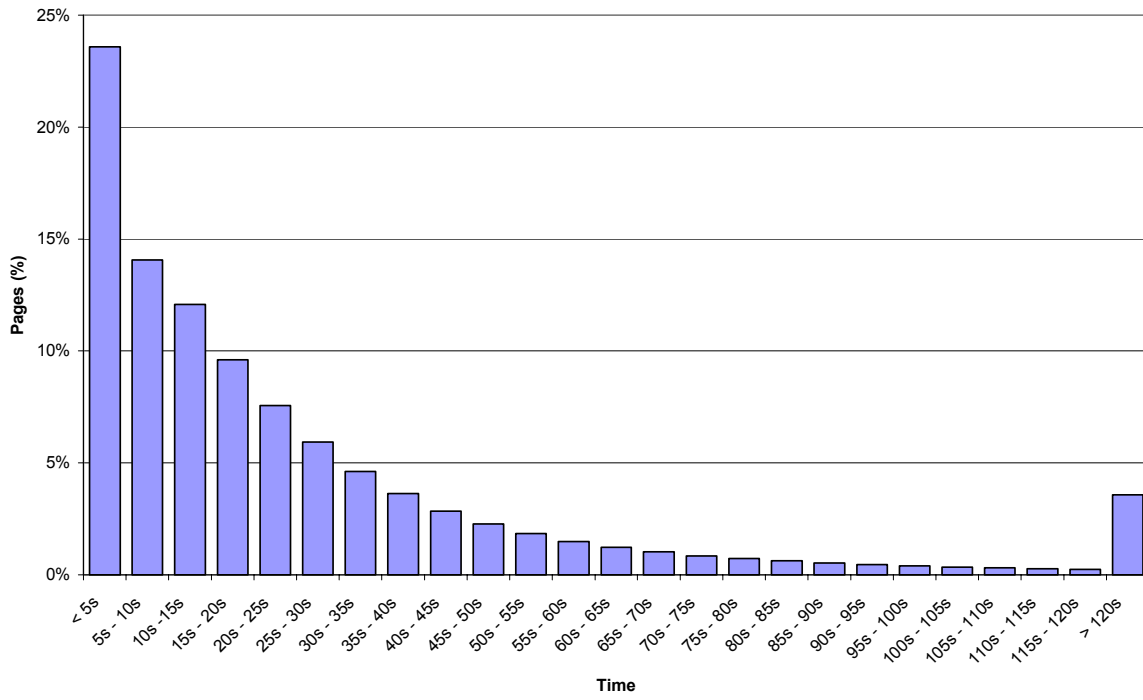
Average Normalized Spending Distribution



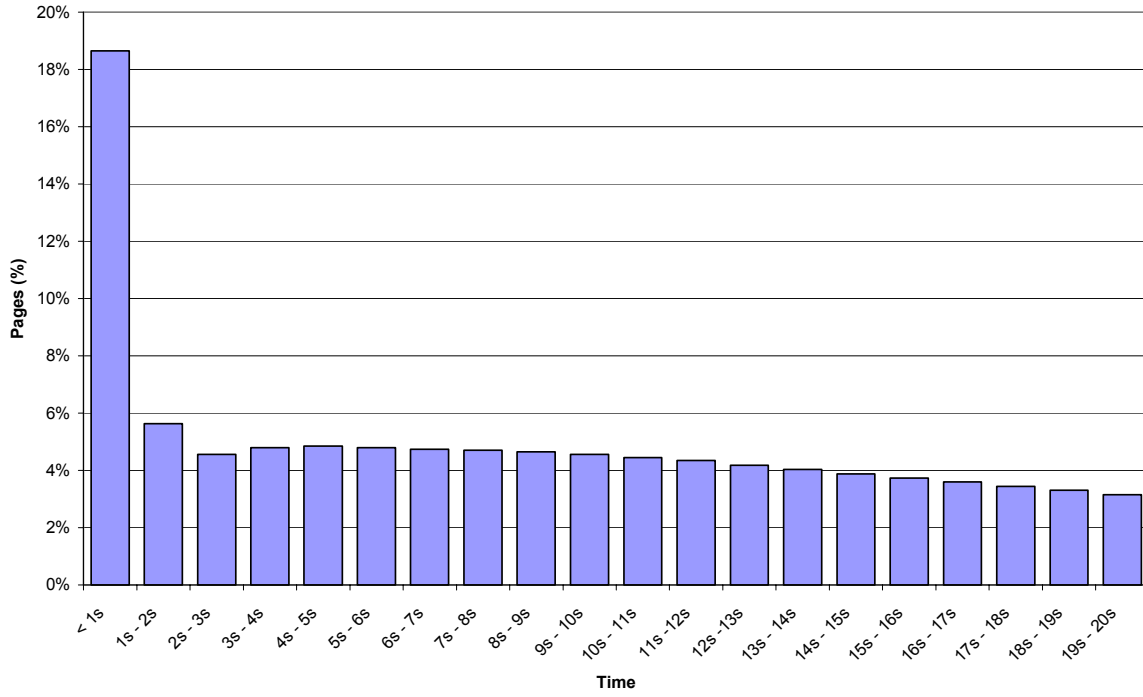
Average Visit to Buy Conversion Rate



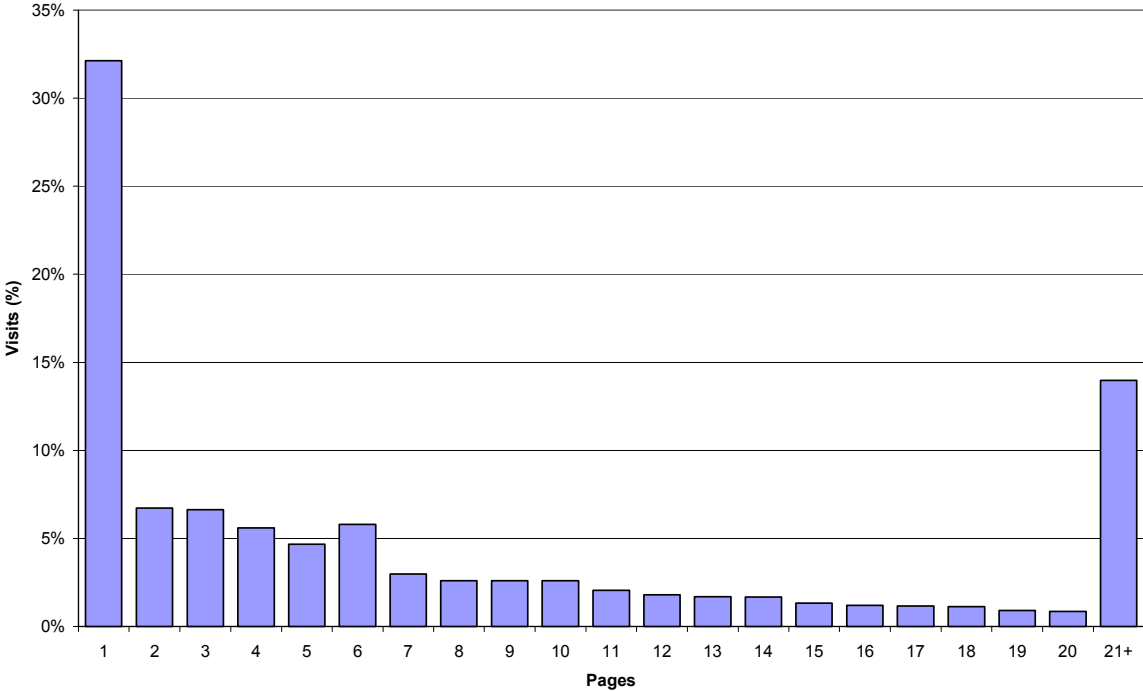
Average Think Time



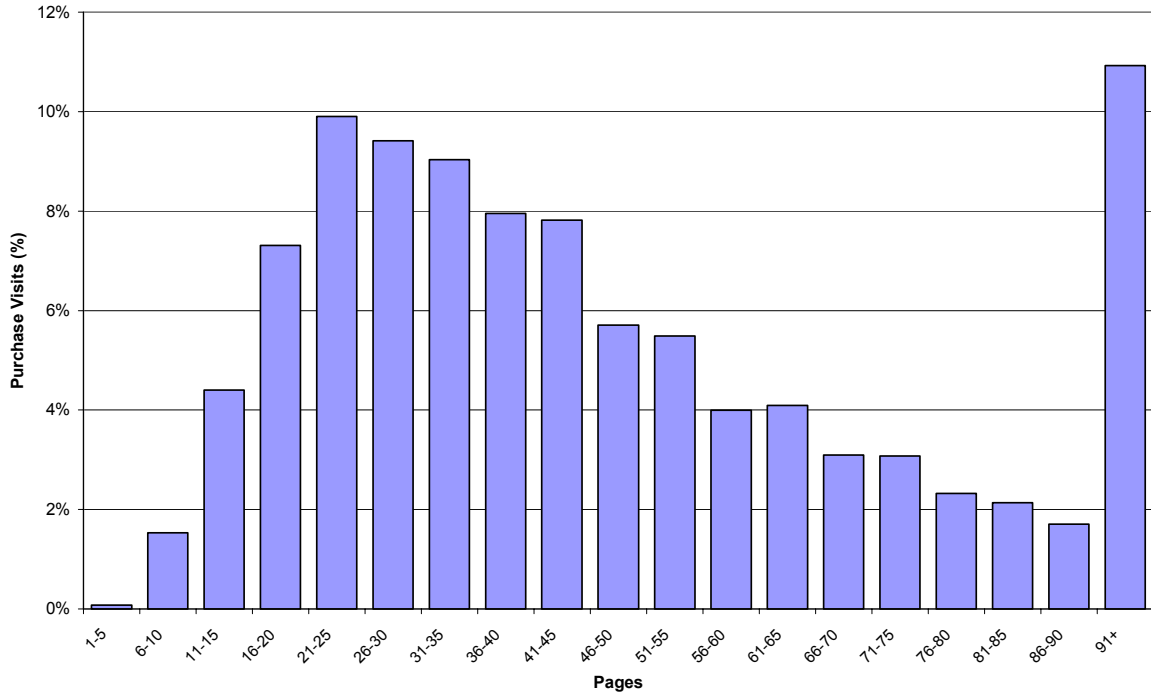
**Average Think Time
(Less than 20s only)**



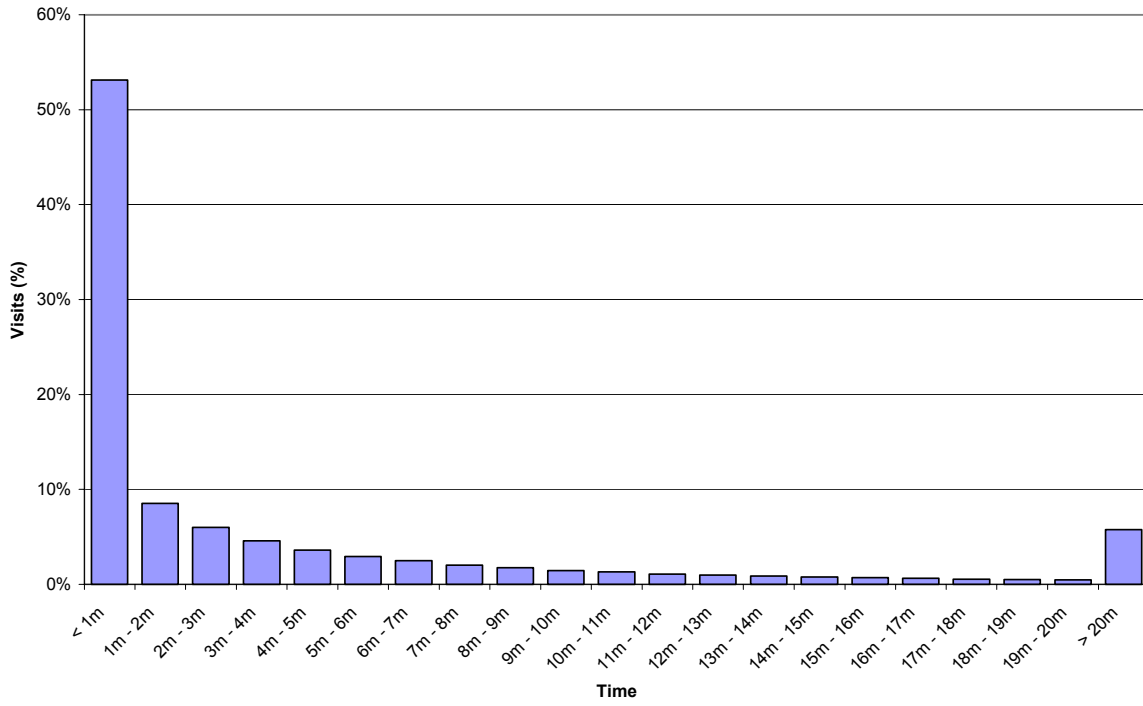
Average Visit Duration (Pages)



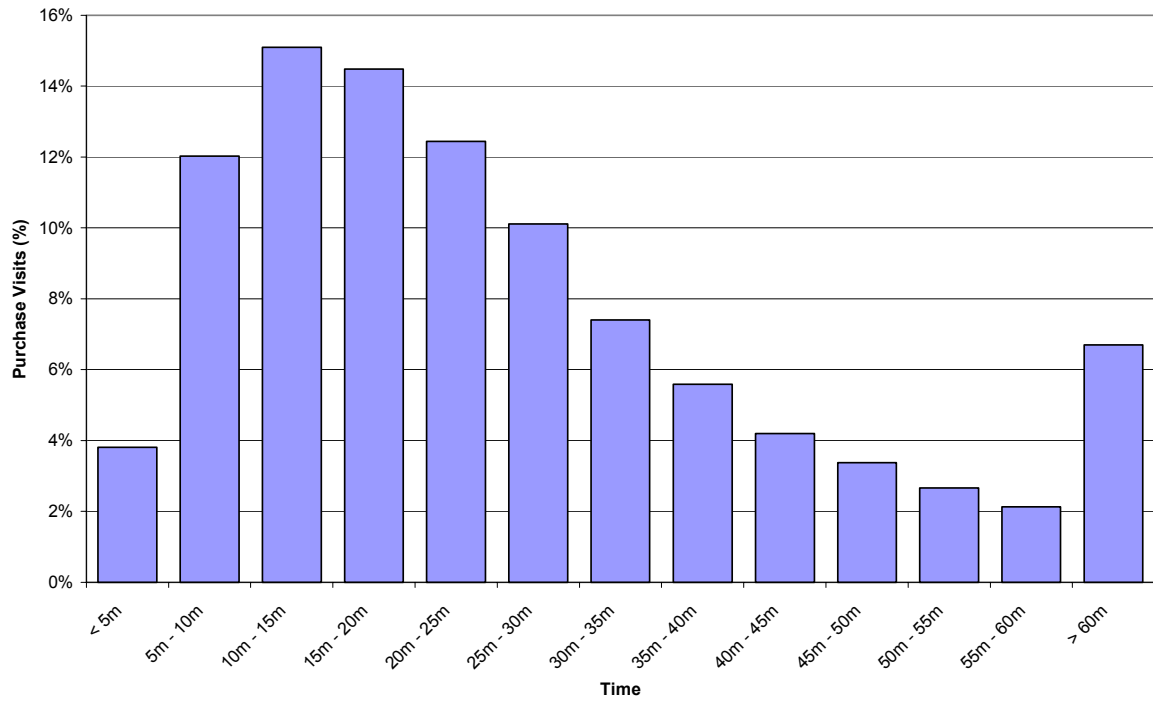
Average Visit Duration (Pages) - Purchase Visits



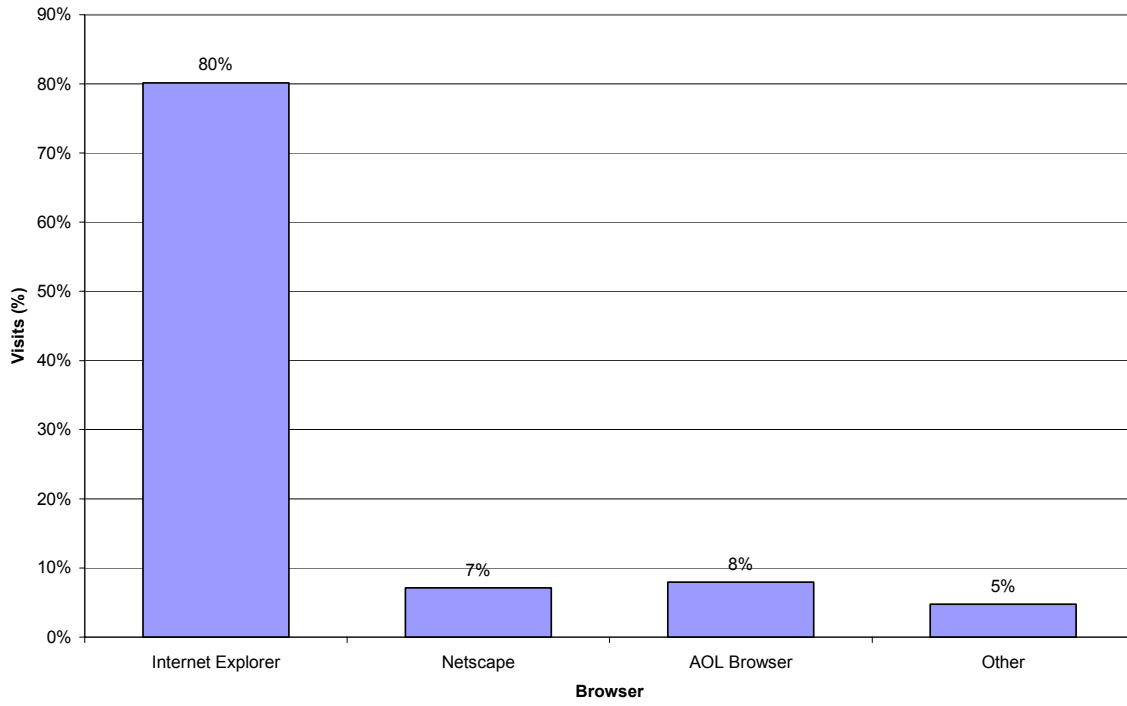
Average Visit Duration (Time)



Average Visit Duration (Time) - Purchase Visits



Average Browser Distribution



4. Comparative Demographics

The following charts illustrate the important demographic characteristics for distinguishing online shoppers from a random sample of the US population.

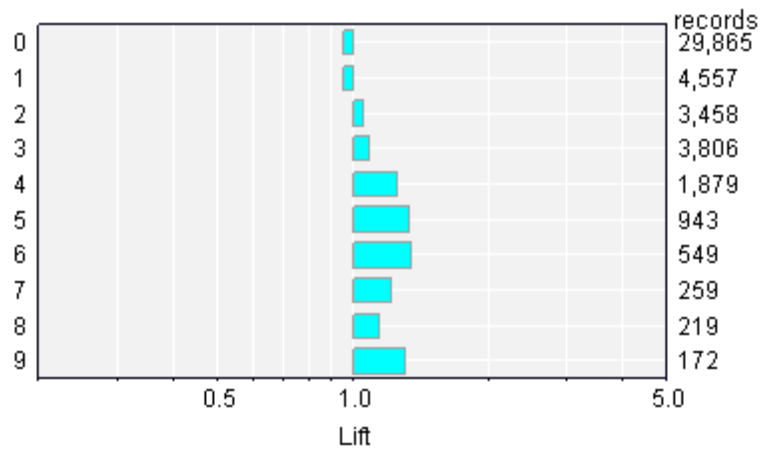
The following Acxiom based demographics are shown below.

Number of Credit Lines
Age
Home Market Value
Household Income
Length of Residence Less Than Five Years
Marital Status
New Car Buyer
Owns Premium Credit Card
Owns Travel and Entertainment Credit Card
Owns Upscale Store Credit Card

The *lift charts* below show the increase or decrease of the probability of being an online shopper based on demographic attributes. This is expressed in bars to the right or left of a baseline of 1.0. Each plot shows the increase in likelihood of being an online shopper given the values of a demographic attribute. The x-axis has its origin at 1 so that bars going to the right have lifts greater than one, while bars going to the left indicate lifts less than one. A value greater than one (to the right) implies that for the given value of the attribute you are more likely to be an online shopper. A value less than one (to the left) implies that for the given value of the attribute you are less likely to be an online shopper.

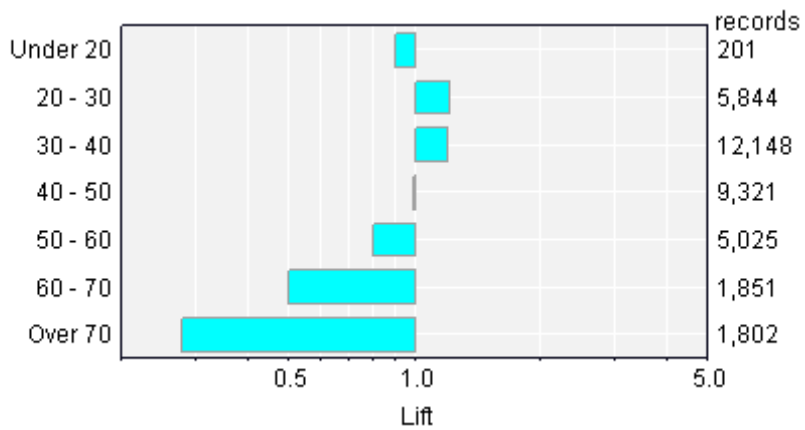
Number of Credit Lines

The more credit lines you have, the more likely you are to be an online shopper.



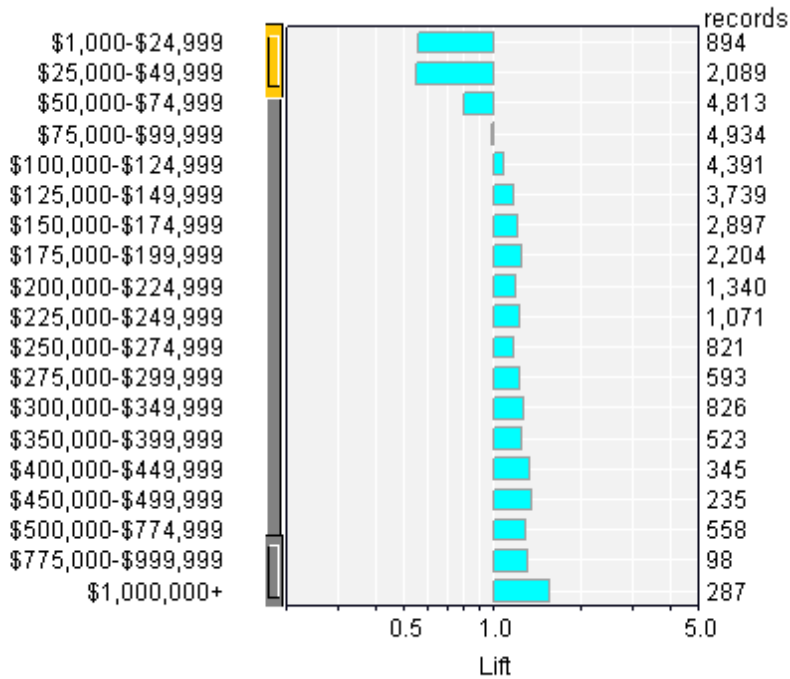
Age

People aged between 20 and 40 are more likely to be online shoppers.



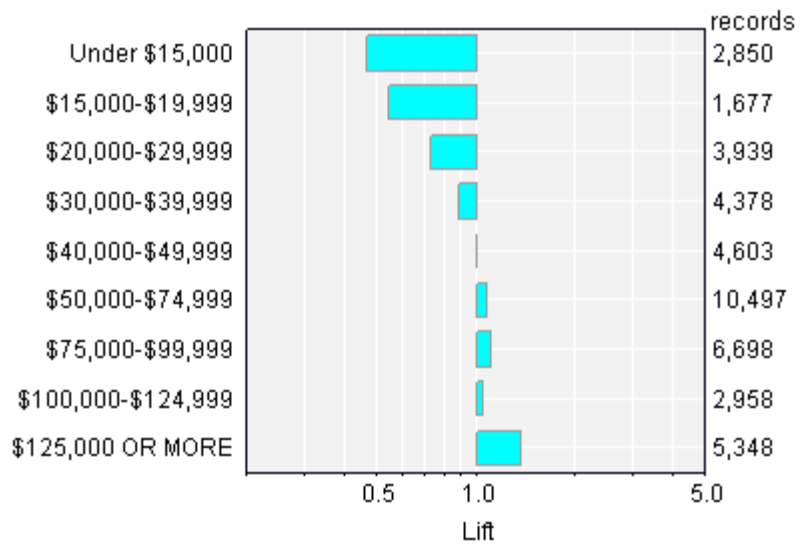
Home Market Value

People with a home market value of more than \$100,000 are more likely to be online shoppers.



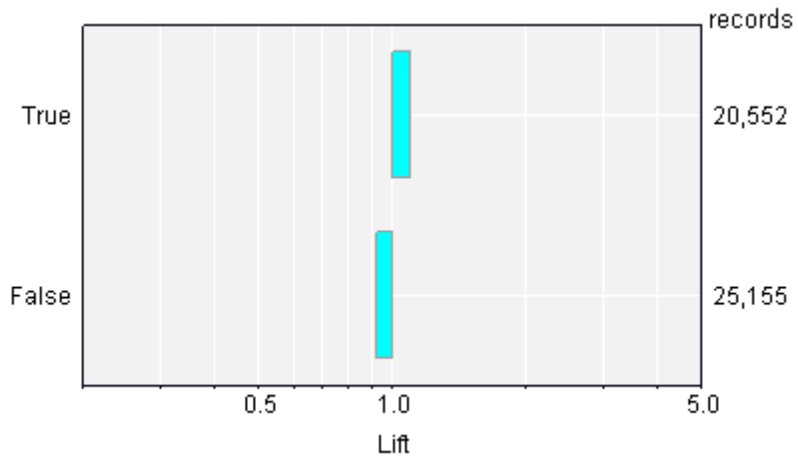
Household Income

People with a household income of more than \$50,000 are more likely to be online shoppers.



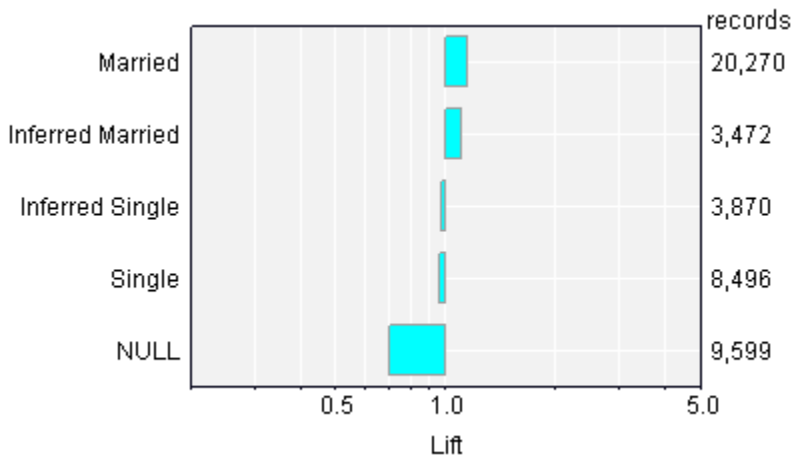
Length of Residence Less Than Five Years

People who have been at their current residence less than five years are more likely to be online shoppers.



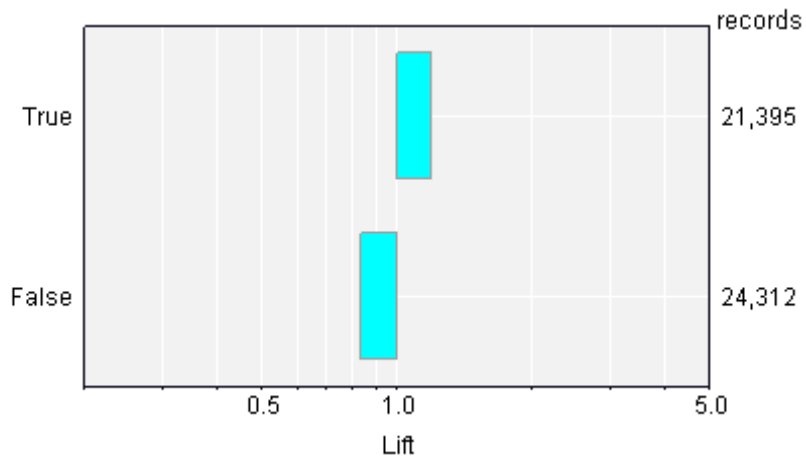
Marital Status

People who are married are more likely to be online shoppers.



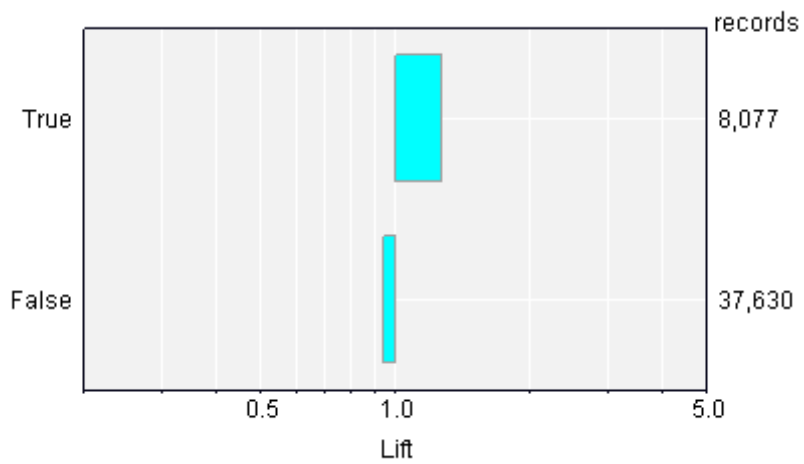
New Car Buyer

People who are new car buyers are more likely to be online shoppers.



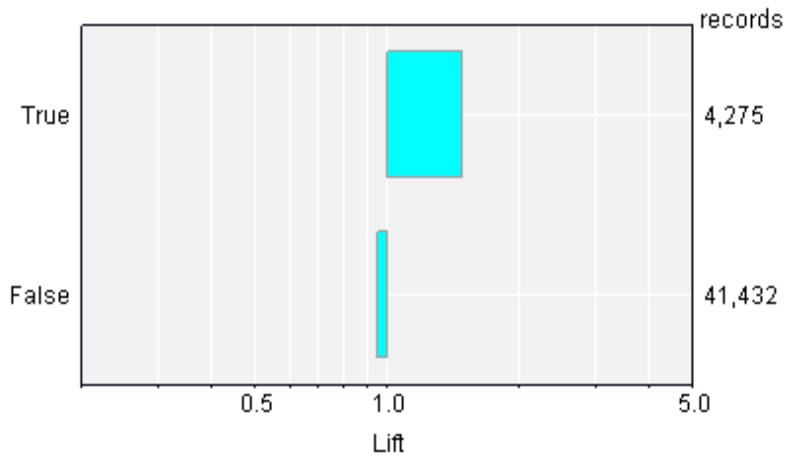
Owns Premium Credit Card

People who own a premium credit card are more likely to be online shoppers.



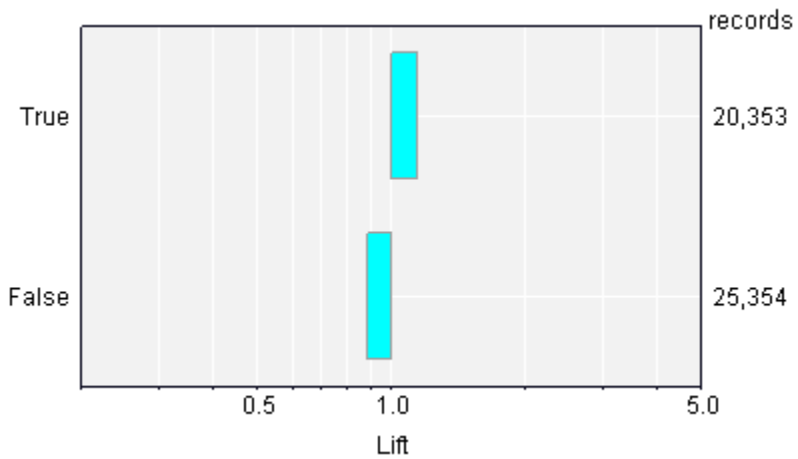
Owns Travel and Entertainment Credit Card

People who own a travel and entertainment credit card (Diners Club or American Express) are more likely to be online shoppers.



Owns Upscale Store Credit Card

People who own an upscale store credit card are more likely to be online shoppers.



5. eMetrics Definitions

All analysis was restricted to the period between 1 November 2000 and 31 December 2000. For eMetrics calculated by week, the period of analysis was reduced to be between 6 November 2000 and 31 December 2000. This period corresponds to eight full weeks of data (Monday to Sunday). One site had clickstream logging disabled until early December. For this site we conducted a reduced study, restricted to the period between 11 December 2000 and 7 January 2001, corresponding to four full weeks of data.

The eMetrics are defined below.

Visits by Week

This is the percentage of the total visits that occurred during each week (Monday to Sunday).

All visits due to robots/crawlers were ignored.

Sales by Week

This is the percentage of the total sales that occurred during each week (Monday to Sunday). Sales is the sum of the 'Order Amount' over all valid orders. A valid order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Visitor to Registered Visitor Conversion Rate

This is the percentage of unique visitors that had not registered at the site before the analysis period that registered at the site during the analysis period. A unique visitor is defined by a unique cookie identifier. A registered visitor is a visitor with a non-null 'Customer ID', identifying them as someone that has either registered or purchased at the site.

Registered Visitor to Buyer Conversion Rate

This is the percentage of registered visitors that had not purchased at the site before the analysis period that purchased at the site during the analysis period. A registered visitor is a visitor with a non-null 'Customer ID', identifying them as someone that has either registered or purchased at the site. A purchase is defined as an order that is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Visitor to Buyer Conversion Rate

This is the product of the visitor to customer conversion rate and the customer to buyer conversion rate.

All visits due to robots/crawlers were ignored.

Visitor to Registered Visitor Conversion Rate by Week

For each week (Monday to Sunday), this is the percentage of unique visitors that had not registered at the site before that week and registered at the site during that week. A unique visitor is defined by a unique cookie identifier. A registered visitor is a visitor with a non-null 'Customer ID', identifying them as someone that has either registered or purchased at the site.

All visits due to robots/crawlers were ignored.

Registered Visitor to Buyer Conversion Rate by Week

For each week (Monday to Sunday), this is the percentage of registered visitors that had not purchased at the site before that week and purchased at the site during that week. A registered visitor is a visitor with a non-null 'Customer ID', identifying them as someone that has either registered or purchased at the site. A purchase is defined as an order that is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Visit to Buy Conversion Rate

This is the percentage of visits that were associated with a valid order. A visit is associated with a valid order if there is an order header with a matching 'Session ID' and that order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Visit to Buy Conversion Rate by Week

For each week (Monday to Sunday), this is the percentage of visits that were associated with a valid order. A visit is associated with a valid order if there is an order header with a matching 'Session ID' and that order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Visitor to Repeat Visitor Conversion Rate

This is the percentage of unique visitors that visited the site for the second time during the analysis period. This was based on visits where the ‘Session Visit Count’ was equal to two. A unique visitor is defined by a unique cookie identifier.

This metric was only calculated for sites that had 1-in-1 clickstream logging enabled for the analysis period.

All visits due to robots/crawlers were ignored.

Average Visits before Registration

This is the average number of previous visits for a visitor at the time of registration (including the visit in which registration occurred). This was based on the ‘Session Visit Count’ for visits where the ‘Account Creation Date’ fell between the ‘Session First Request Date’ and ‘Session Last Request Date’.

This metric was only calculated for sites that had 1-in-1 clickstream logging enabled for the analysis period.

All visits due to robots/crawlers were ignored.

Normalized Spending Distribution

This is the distribution of customers’ average order amounts. To allow comparison across different sites, we normalized the average order amount using:

$$\text{NORMALIZED AMOUNT} = \frac{\text{CUSTOMER AVERAGE ORDER AMOUNT} - \text{MINIMUM ORDER AMOUNT}}{\text{MAXIMUM ORDER AMOUNT} - \text{MINIMUM ORDER AMOUNT}}$$

A value of 0 corresponds to a customer spending the minimum amount. A value of 1 corresponds to a customer spending the maximum amount. A value of 0.5 corresponds to a customer spending the average amount. To avoid outliers skewing the results, we removed the top 2.5% of orders by amount.

All visits due to robots/crawlers were ignored.

Shopping Cart Abandonment Rate

This is the percentage of shopping carts that were abandoned.

This metric was only calculated for sites running v3.1 or later with cart business event logging enabled.

All visits due to robots/crawlers were ignored.

Percentage of Visits Using Search

This is the percentage of visits that used search.

This metric was calculated exactly for sites running v3.1 or later with search business event logging enabled. Otherwise, it was calculated approximately as the percentage of visits that included a page view of the search page or search results. In this case, the JSP templates containing search were identified by manual inspection of the site. The resulting metric is only an approximation because it is possible that a search page could have been viewed without a search actually being performed.

All visits due to robots/crawlers were ignored.

Percentage of Failed Searches

This is the percentage of searches which returned zero results.

This metric was only calculated for sites running v3.1 or later with search business event logging enabled.

All visits due to robots/crawlers were ignored.

Percentage of Visitors Reading the Privacy Statement

This is the percentage of unique visitors that visited the page containing the privacy statement on any of their visits. The JSP template containing the privacy statement was identified by manual inspection of the site. A unique visitor is defined by a unique cookie identifier.

All visits due to robots/crawlers were ignored.

Average Visit Duration (Page Views)

This is the average visit duration in terms of the number of pages viewed. The visit duration is the number of page views for the visit.

All visits due to robots/crawlers were ignored.

Average Purchasing Visit Duration (Page Views)

This is the average visit duration in terms of the number of pages viewed, restricted to visits that are associated with a valid order. The visit duration is the number of page views for the visit. A visit is associated with a valid order if there is an order header with a matching 'Session ID' and that order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Visit Duration (Page Views) Distribution

This is the distribution of the visit duration in terms of the number of pages viewed. The visit duration is the number of page views for the visit.

All visits due to robots/crawlers were ignored.

Purchasing Visit Duration (Page Views) Distribution

This is the distribution of the visit duration in terms of the number of pages viewed, restricted to visits that are associated with a valid order. The visit duration is the number of page views for the visit. A visit is associated with a valid order if there is an order header with a matching 'Session ID' and that order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers were ignored.

Average Visit Duration (Time)

This is the average visit duration in terms of time. The visit duration is calculated as the difference between the 'First Request Date' and 'Last Request Date' columns for the visit.

Visits with only a single page view were ignored.

All visits due to robots/crawlers were ignored.

Average Purchasing Visit Duration (Time)

This is the average visit duration in terms of time, restricted to visits that are associated with a valid order. The visit duration is calculated as the difference between the 'First Request Date' and 'Last Request Date' columns for the visit. A visit is associated with a valid order if there is an order header with a matching 'Session ID' and that order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers are ignored.

Visit Duration (Time) Distribution

This is the distribution of the visit duration in terms of time. The visit duration is calculated as the difference between the 'First Request Date' and 'Last Request Date' columns for the visit.

Visits with only a single page view are ignored.

All visits due to robots/crawlers are ignored.

Purchasing Visit Duration (Time) Distribution

This is the distribution of the visit duration in terms of time, restricted to visits that are associated with a valid order. The visit duration is calculated as the difference between the 'First Request Date' and 'Last Request Date' columns for the visit. A visit is associated with a valid order if there is an order header with a matching 'Session ID' and that order is not a return, has a status of complete, submitted, partially shipped or shipped, and has an order amount greater than zero.

All visits due to robots/crawlers are ignored.

Average Think Time

This is the average amount of time that passed between consecutive page views. The amount of time between consecutive page views is calculated as the difference between the 'Request Date' on page view X and the 'Request Date' on page view $(X + 1)$.

The last page view of a visit is always ignored.

All page views due to robots/crawlers are ignored.

Think Time Distribution

This is the distribution of the amount of time that passed between consecutive page views. The amount of time between consecutive page views is calculated as the difference between the 'Request Date' on page view X and the 'Request Date' on page view $(X + 1)$.

The last page view of a visit is always ignored.

All page views due to robots/crawlers are ignored.

Average Home Page JSP Processing Time

This is the average amount of time spent in processing (compilation and execution) of the home JSP page to generate HTML. This is calculated as the average of the 'Compile Time' column over all page views of the home page. The JSP template for the home page was identified by manual inspection of the site.

To avoid outliers skewing the average we ignored any compile time greater than 10 minutes.

All page views due to robots/crawlers are ignored.

Average JSP Processing Time

This is the average amount of time spent in processing (compilation and execution) a JSP page to generate HTML. This is calculated as the average of the 'Compile Time' column over all page views.

To avoid outliers skewing the average we ignored any compile time greater than 10 minutes.

All page views due to robots/crawlers are ignored.

Percentage of Visits due to Robots/Crawlers

This is the percentage of visits that were marked (during preprocessing) as being generated by robots or crawlers.

Visit Browser Distribution

This is the distribution of user agents visiting the site based on the number of visits by each type of user agent. The user agents were grouped into five categories: Internet Explorer, Netscape Navigator, AOL Browser, Robots/Crawlers and Other.

6. Acknowledgements

We thank Acxiom Corporation for providing the syndicated data used to generate comparative site demographics. We thank the members of the Blue Martini data mining team.