

# Clickstream Analysis to Forecast Online Behavior

Ram Ganeshan

## CLICKSTREAM ANALYTICS

A *clickstream* is a sequence of clicks a customer goes through on a website before he or she exits the site. When the website is an online retailer, the clickstream gives an insight into how the customer navigates the online store. A potential customer can have multiple entry points into the site – they could have searched for a relevant product, responded to a promotion, or visited the site with an intent to research, compare, and perhaps purchase the products the store has to offer.

Once the customer has entered the site, he or she could visit multiple pages, maybe even add products to the “Shopping Cart,” and eventually exit the site with or without making a purchase. Widely available web-analytics applications -- such as Google Analytics<sup>1</sup> -- make it easy and relatively cheap to capture a customer’s clickstream.

## WEB ANALYTICS

Once installed, web-analytics applications typically work by creating an entry in a *weblog* each time a visitor – a potential customer – visits a web page. This entry contains information about the customer’s IP address, identity, the page visited,

time spent, where they came from, and where they exited. Using these weblogs, it is possible to identify the “path” taken by this visitor through the web site.

Analysis of these visitor paths combined with ecommerce transactional information to cull useful insights into consumer behavior is referred to as *click-stream analysis*.

## **CLICKSTREAM ANALYSIS**

Clickstream analysis has significant impact for online retailers. With it, they can recreate the experience of the neighborhood store – understand their customer, provide useful and relevant information; and make the purchasing experience a pleasant one.

With the availability of large amounts of data from the clickstream, the forecaster’s role – at least in firms with a significant online presence -- has also been expanded. The traditional role of generating item-level forecasts has to be integrated with information on customer browsing behavior. This has the potential advantage of not just increasing revenue, but the added benefit of understanding the customer’s intent – every forecaster’s holy grail.

My intention in this article is to show, based on my experiences with online retailers, how one can use clickstream data – in this case collected from Google Analytics – to explore three broad questions:

- How do we optimize the path taken by the customer in the web site? The primary idea is to explore and analyze how *browsing* can ultimately lead to *buying*.
- How can a retailer establish associations between products? Commonly called *Market Basket Analysis*, this analysis can inform forecasting, promotions, and replenishment.
- Finally, how can an online retailer modify its web site based on the click stream to make the shopping experience a more fruitful one?

While there are multiple ways and several layers of analysis involved to explore these questions, I show how readily available data and visualization tools from Google Analytics can get any forecaster at an online retailer quickly started on clickstream analysis.

## **CUSTOMER PATH OPTIMIZATION**

In Google Analytics, a *visitor* is someone who has entered a web site. A *conversion* occurs when the visitor has completed a pre-determined *goal* that is set by the web site owner. For the online retailer, a visitor<sup>2</sup> is a potential

customer, somebody who has entered the web shop front. Conversion implies that this potential customer has purchased products from this store.

## CONVERSION RATES

Exhibit 1 gives the comparison between visitors and conversion rates for an online retailer I have been working with (the data is scaled to protect confidentiality). The visits and the conversions are usually highly correlated, certainly for this retailer. In Exhibit 1, the conversion rates are 2-4% with a peak of about 8% during the holiday season (driven largely by promotions). This is typical of many online retailers. The way to increase revenue is to drive more traffic to the web site; and generate more conversions from these visitors.

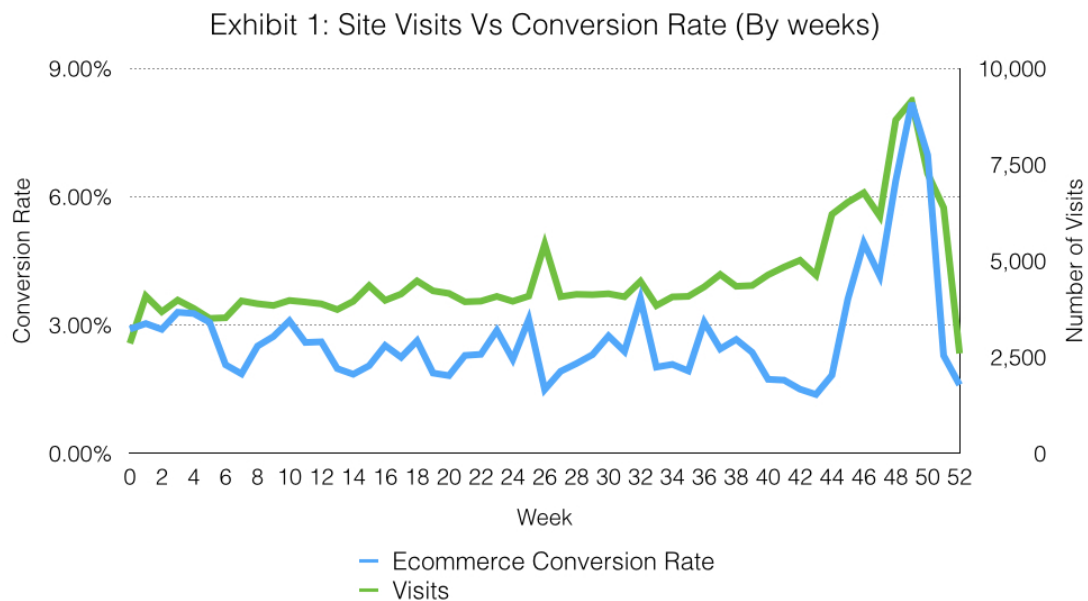


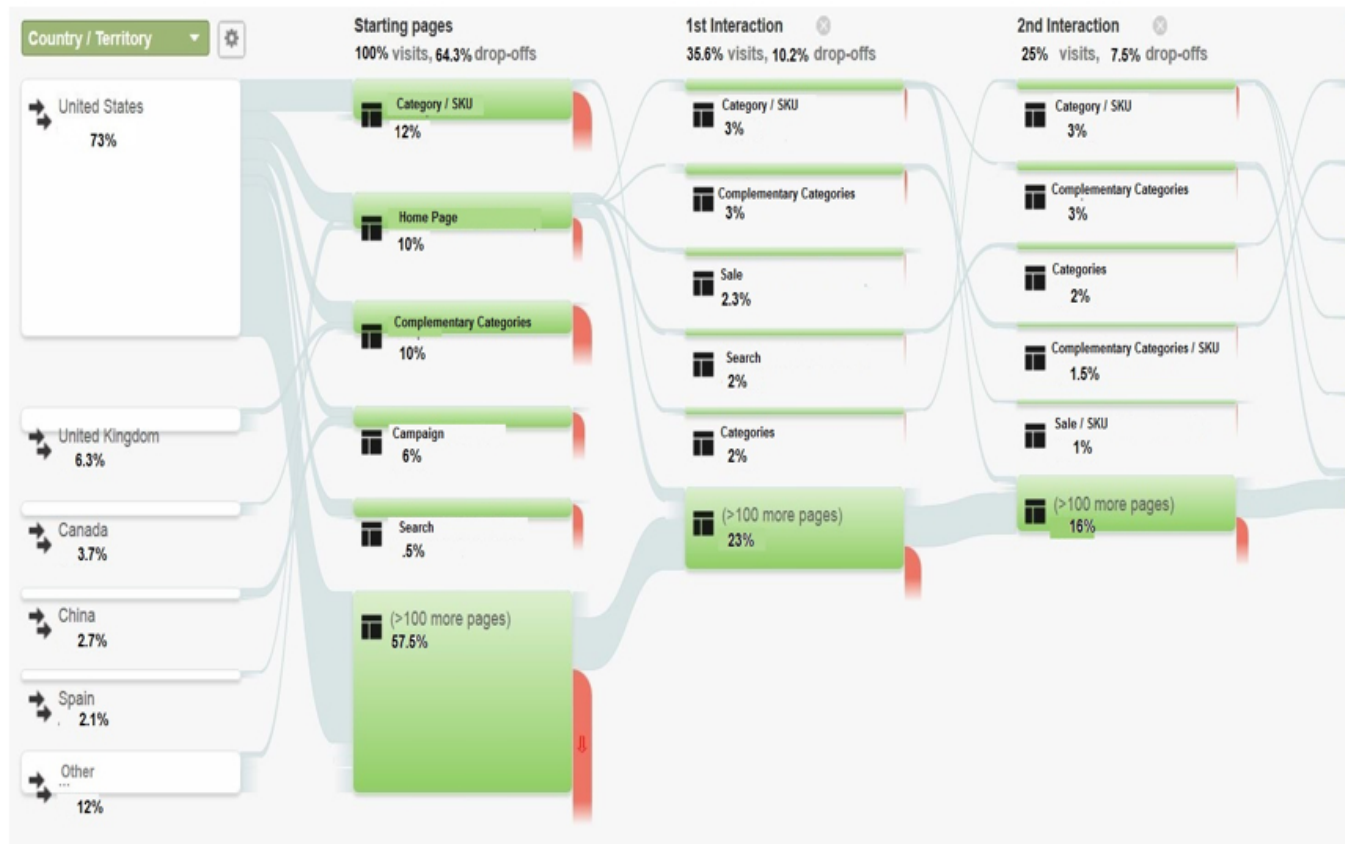
Exhibit 2 shows a *visitor flow chart* readily generated in Google Analytics. The chart reveals where the customers come from (in this case most are from the USA) as well as the paths, on average, taken by these customers. After landing on a starting page, about 35.6% of visitors proceed (64.3% leave the site) to navigate through product categories (we have for reasons of confidentiality hidden the actual web site paths). For example, one traditional path to adding products to carts is:

Unites States -> Home Pages ->Categories ->Category SKUs ->Add to cart

The chart shows that it takes this customer segment at least 3 clicks on the web site before a product is added to the cart.

However, there are numerous paths (some that loop without ending in a cart) and such paths can give insights into whether the customer is actually navigating the site as intended. The retailer can than test different configurations of its web site to enhance visitor-browsing behavior. Paths can also be visualized for different customer segments – especially useful if the customer is identified by a login – to analyze if these segments are navigating the web site as intended.

## Exhibit 2: Visitor Path



The entrance paths can also provide insights on which campaigns or promotions (Google search, paid search, catalog, email promotions, etc.) are attracting customers (Google Analytics allows the website owner to segment entrance paths along several dimensions) and those that are actually resulting in a

conversion. Resources can then be efficiently allocated among these marketing channels.

### ***GOAL FUNNELS and ABANDONMENTS***

Google Analytics also allows the website owner to create a *goal funnel* – a tool to visualize the various steps the customer takes complete a purchase. Exhibit 3 shows the *goal funnel* for an online retailer. Once customers have added products to their shopping cart and are ready to check out, they have to go through three steps: they need to fill out relevant information in the “Bill to/Ship to” page, then proceed to shipping and choose the appropriate shipping method, and finally use a credit card or bank account to pay for their purchase. Of the customers who enter the “Bill to /Ship to” page, 18.4% simply abandon their carts. Of those that actually filled in all the information, 15.2% decide to leave when they get to the shipping page. Finally, 8.46% leave when they are ready to pay. All these abandoning customers clearly have intent, i.e., they filled cart and came to checkout process – customers that could have generated revenue.

The online retailer can use the goal funnel (again this funnel can be filtered by various customer segments) to gather insight into why the cart was abandoned. Perhaps the “Bill to /Ship to” page is too onerous or perhaps the shipping costs are too high or their preferred credit card option was not available. If the identity of this customer can be established (through “logins” or in some cases via

“cookies”) it may be possible to reach out to them “offline” to follow up on the missed sale. Such funnels, at the very least, enable retailers to streamline the checkout process to reduce cart abandonment.

Exhibit 3: The Goal Funnel



## MARKET BASKET ANALYSIS

Market Basket Analysis allow retailers to quickly and easily examine the size, contents, and value of their customer’s market basket and identify what products (or web sites) tend to be purchased (visited) together. In the most general sense, these purchases can be in the same cart (for example chips and salsa purchased



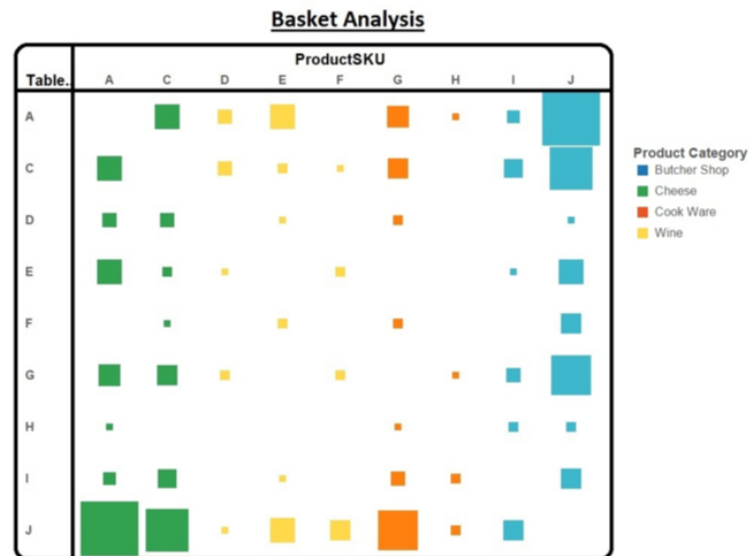
together), or purchased sequentially, such as a home and then furniture at a later date.

-Google Analytics can track ecommerce transactions for any web site. Product association can be visualized by counting the number of times any two products appear together in a cart (or for the same customer, between products in different carts over time). Exhibit 4 shows the top 10 selling SKUs of a mid-size online retailer in the Food and Cookware space. Google Analytics data was downloaded to Tableau, a data visualization tool, to generate this visual<sup>3</sup>.

Looking along the first row, every time the customer buys SKU A (in the Cheese category); he/she has the highest likelihood of buying product J (in the Butcher Shop). SKUs G (in Cookware) and E (in Wine) have also a high likelihood of being purchased. The size of the squares represents the magnitude of the

association and color represents the product categories.

Exhibit 4



Market basket analysis can be useful in several ways. First, associated SKUs can be merchandized together. In a traditional store, we have all seen displays of eggs next to milk; or travel clocks next to luggage. For online retailers, this proximity could mean displaying associated products together – for example Amazon.com’s *frequently bought together* information right below the product description. In our example, when the customer lands on a product page for A, products D and J can be displayed on the same page to increase cross-sells. This will make the user experience better and perhaps lead to more sales.

Second, from an operations perspective, planners can forecast associated products together; and synchronize the timing of the replenishment of these products. Third, by analyzing the baskets of loyal customers over time, the retailers can make targeted promotions to increase margins. For example, a customer segment can be enticed to enter the store for a promotion on a certain product (say SKU A in our example) but eventually also buy associated products (SKUs G and D) at full price, the retailer has simultaneously increased traffic, engagement, and likely, the revenue.

## **IMPROVING THE SHOPPING EXPERIENCE**

Optimizing visitor paths and market basket analysis go a long way toward making the shopping experience useful and relevant. However Google Analytics also has tools that help analyze if the web site is in fact providing the customer useful and relevant information in order to make purchases.

Exhibit 5 shows the top 15 search terms potential customers have used, when they are in the online store (The data are from an online retailer, but scaled).

These are customers are looking for something specific. For example, search term #4 was searched 1076 times with only 6 of those yielded a conversion, with an average value of \$148.13, a rate of 0.56%. Term #8, on the other hand, was searched 841 times but yielded only 2 transactions (a 0.24% conversion rate, each with an average value of \$44.20. These customers are either not finding

the items they seek or the search is yielding results that are not useful. Perhaps item #4 was too expensive.

Exhibit 5: Search Terms and Conversions

Search Term	Total Unique Searches	Transactions	Avg. Order Value	Ecommerce Conversion Rate
1	1337	54	\$64.28	4.04%
2	1187	100	\$65.81	8.42%
3	1113	13	\$72.64	1.12%
4	1076	6	\$148.13	0.56%
5	1065	12	\$82.44	1.13%
6	1047	10	\$36.59	0.91%
7	949	108	\$57.99	11.39%
8	841	2	\$44.20	0.24%
9	787	10	\$98.85	1.21%
10	767	21	\$29.42	2.67%
11	677	28	\$60.01	4.07%
12	668	58	\$77.87	8.61%
13	631	41	\$61.00	6.42%
14	627	50	\$62.72	7.97%
15	622	54	\$72.39	8.60%

The retailer can test causes for such poor conversions by what is called A/B testing – customers are randomly assigned one of two web pages and the retailer can then look at the conversion rates to redesign the web site experience.

## CHALLENGES

This article highlights how easily available and relatively cheap tools such as Google Analytics can provide insights into customer behavior in the clickstream. However, analyzing the clickstream brings with it a host of new challenges. First,

the clickstream generates an enormous amount of data – even for a mid-size retailer the number of entries in the weblog can quickly run into the millions. While Google Analytics' tools help collate the data via simple algorithms, significant resources may be needed to fully leverage the clickstream. For example, investment in software and hardware infrastructure to store and interpret the vast amount of data may be warranted; and in many cases, it may necessitate hiring personnel who are competent in handling and interpreting such vast amounts of data. Second, online retailers, and indeed any users of such clickstream data, must have a strategy to integrate clickstream data into traditional forecasting methods – i.e., clearly define how it impacts the sales forecasting and operations functions in the firm<sup>4</sup>. However, the results can be rewarding, and in my experience, provide significant returns on investment.

---

<sup>1</sup> For detailed information on Google Analytics see: <http://analytics.google.com>

<sup>2</sup> In a recent article in this journal Schmidt and Vosen show how using another Google tool, “Google trends” which shows industry trends, a forecaster can predict consumption patterns. See Schmidt, T. and Vosen, S. (2013). Forecasting Consumer Purchases Using Google Trends, *Foresight*, Summer, 38-41.

<sup>3</sup> <http://kb.tableausoftware.com/articles/knowledgebase/market-basket-analysis>

<sup>4</sup> Patrick McSharry gives a great overview of “Stream Analysis” (the clickstream is a type of stream analysis) and its implementation challenges. Much of execution challenges he describes are directly applicable to clickstream analysis. See McSharry, P. (2012). Stream Analytics for Forecasting, *Foresight*, Winter, 7-12.