

Title: Patterns of online grocery purchase: first and repeat buyers

Authors

Wasimakram Binnal
Department of Management Studies
Indian Institute of Science
Bengaluru – 560012, India.
wasimakramb@iisc.ac.in

Anjula Gurtoo
Department of Management Studies
Indian Institute of Science
Bengaluru – 560012, India.
anjula@mgmt.iisc.ernet.in

Abstract

The paper aims for better understanding of online consumer decision making in low involvement products by exploring the differences between first and repeat buyers. Collaboration with leading online grocery retailer in India was formalized for data. The retailer gets more than a million clicks each day. The following data was shared by the retailer and forms the base for the analysis - server weblog information like purchase-orders, browsing patterns, timestamp, session details, ip-address, device used, navigation pages and browsers used. The tab-delimited data is for 20 days in total. Data cleaning is done with a focus on answering the research questions. Multimethod research methodology is used for data analysis. Descriptive statistics and CART decision trees techniques are used. The findings provide insight into online consumer behaviour in low involvement products. There is no difference in average time spent on webpage between first and repeat buyers and there is a significant difference between first and repeat buyers in order value. First time consumers exhibit hesitation in using navigational choices, specifically product description and product search. The first-time consumers exhibit no clear path but have higher inclination towards browsing of products. Fruit-vegetables and grocery staples is the most visited category item for both.

Keywords

Online purchase; first-time buyers; repeat buyers; weblogs; grocery categories; navigational choices; source websites.

Introduction

Internet has become the primary source of information for most consumers (Zha, Li and Yan, 2013). Online retailers provide higher quality information and more information about the product (Peterson et al., 1997; Wolfinbarger and Gilly, 2001; Kolesar and Galbraith, 2000). This makes the whole process of buying online complex as consumers are exposed to a wide range of information. Consequently, consumers look for efficient and heuristic decisions. While access to a large amount of information creates a perception of control in stressful situations (Parra and Ruiz, 2009), information overload causes confusion, stress, inefficiency and unfavourable decisions. If consumers lose perception of control, that results in consumer ceasing browsing (Korgaonkar and Wolin, 1999).

First time buyers are not completely familiar with the websites features and hence a learning curve is seen before utilization of all the web features optimally. Post the learning curve, repeat buyers buy the same products/services from the same vendor due to their previous shopping experiences (Kuo, Hu and Yang, 2013; Wendy, and Fader, 2004). McCarthy, Sullivan and Reilly (1999) indicate that first time buyers have higher inclination towards innovativeness, especially for unusual or foreign food products. Therefore, price and quality of branded products can be used to forecast first and repeat purchase probability (Baidya and Ghosh, 2014; Moe and Fader, 2001). Repeat purchase buyers are seen to have a fairly focused navigation path (Wendy, 2003).

Repeat purchase orders reflect a consumer's belief about the service and the quality of that particular website and products. Consumers are highly inclined to purchase from online retailers with a past sales record, and this is independent of the purchased items (Ye and Fang, 2013). However, there are no studies to the best of our knowledge that explore first time online grocery purchasers versus repeat purchasers.

Methodology

The raw data is cleaned, preprocessed and prepared for measurement. As research objectives are diverse, the study employs multi-method data mining for exploring the objectives. Multi-method approach offsets any discrepancies and minimizes limitations (Maxwell and Loomis, 2003). Increasingly research is getting inter-disciplinary, complex and dynamic, which requires a shift from the traditional methodological approaches to multi method approaches (Johnson and Onwuegbuzie, 2004).

Weblogs dataset provides information related to consumer behaviour (Poel and Buckinx, 2005). Clickstream data provides a trace of how products were added/removed from the basket, navigational choices taken by consumer and time spent in each stage of online shopping (Bucklin et al., 2002). A collaboration with a leading online grocery store was formalized in March 2015, with a Non-disclosure-agreement. The company provided raw server weblogs for this thesis. The company has centers in the metro cities of Hyderabad, Bangalore, Chennai and Mumbai, and is providing online grocery delivery services since 2011. The company gets more than one million clicks per day. The data procured from the retailer is a raw log data file that is tab delimited, with first set of nine days' data from August 01-09, 2014, and the second set of eleven days of data from January 20-31, 2015. The Server weblogs are Tab delimited with 10 labels, with details on navigation, ip-address, session-details, device used, browsers used, web-links, server-request, source-navigation, unique-id and date. The data has messy structure with varying length and alphanumeric. Each string has variation in the length. Timestamp is recorded with information of device details and operating system details as well.

Astudillo et al. (2014) highlights data mining techniques becoming more useful and more suitable to discover and understand unknown consumer patterns. They suggest “The use of Web

log data permits an understanding of user behavior, contains information about user access and may show potential patterns in their behavior. The process of data cleaning and preparation follows Wickham (2011) who suggested the split-apply-combine strategy. This study uses the following steps for data cleaning and preparation. Step.1. Raw data filtered by removing unwanted values, columns and rows. Step.2. Split data into independent strings. Step.3. Grouping of records with same IP-address (includes Unique-id). Step.4. Generating cleaned data list.

Data Analysis and Results

Kahn and Lehmann (1991) suggest large assortments tax consumers' cognitive abilities. Nicholas et al.'s (2005) study on scholar online searches found only 3% of the users choose to search across different categories. However, in online grocery it may be different. Online grocery stores have large assortments and each items influence is hard to analyse, though behaviour across top items can be analysed.

For 20 days of data, there were more than a million clicks each day with 638862 unique visits with a transition to purchase rate of approximate 10%. Data was prepared separately for each day. A rank value was given based on each session visits. The rank value was aggregated for each day. Result of top items visited was generated by merging 20 days of item ranks. The result in Table 6.5 represents the top 10 items visited by consumers.

Table 1: Top 10 items visited

Sl	Grocery Items	Number_days_on top
1	basmati-rice	18
2	potato-onion-tomato	18
3	sunflower-oils	18
4	cheese	17
5	fish	17
6	bread	13
7	cut-fruits-vegetables	13
8	leafy-vegetables	13
9	fresh-chicken	12

10	bread	12
----	-------	----

As seen basmati-rice is the most prevalent top item. This is expected because data is from the four southern metro cities of India. The existence of cheese within the top item is peculiar as it is not a common item in traditional Indian cuisine. This indicates a lifestyle change influenced by the metro or international culture. Bread and cheese are the top ten items found in a study conducted by Hulten and Vanyushyn (2011) across France and Sweden.

Sondhi (2014) found Indian consumers showing keen interest in organic products and being highly conscious about food choices while shopping groceries. We find the same in our results. Both fruits-vegetables and grocery-staples categories have organic assortments, which the consumers have purchased. While Indian consumption pattern is different from other countries, consumers in urban areas are more globalized. Which is why the top items purchased in Greece, France and Sweden are similar to the top items purchased in India.

Repeat Purchase and First Purchase

This subsection builds a decision model for recognizing various navigational choices and source website choices for repeat and first-time buyers. Consumer patterns for navigation are important indicator of their perception and attitudes (Chu, Arce-Urriza, Cebollada-Calvo and Chintagunta, 2010). CART decision tree is employed and descriptive statistics for the same is provided in Table 2.

Table 2: Conditional inference tree for product categories

Conditional inference tree with seven terminal nodes				
Number of observations: 24457				
Node sl.	Details variable labels	Criterion	Statistics	Weights
1	Submenus (cl), promo, shopping-list, special/seasonal (sp)	1	319.057	
2 to 3	Coupons, retailwebsite.com	1	77.593	
3 to 4	beverages, fruits-vegetables, imported-gourmet	1	37.145	
4 to 5	promo, shopping-list	0.98	10.886	
5 to 6	shopping-list	0.98	7.469	

6	Leaf (end) node			9 ***
5 to 7	promo			
7	Leaf (end) node			132 ***
4 to 8	Submenus (cl), special/seasonal (sp)			
8	Leaf (end) node			2341***
3 to 9	branded-foods, bread-dairy-eggs, grocery-staples, household, meat, personal-care			
9	Leaf (end) node			3592***
2to10	Bing, Google, Other			
10	Leaf (end) node			97 ***
1 to 11	basket, view/review-orders (co), green-basket, Offers, products in basket(pb), product-categories mainpage (pc), product description (pd), product-search (ps)			
11to12	beverages, household, imported-gourmet	0.975	20.61	
12	Leaf (end) node			3425***
11 to 13	branded-foods, bread-dairy-eggs, fruits-vegetables, grocery-staples, meat, personal-care			
13	Leaf (end) node			14861**

As seen in Figure 1 there are seven leaf nodes with details of the nodes provided in Table

2. Node six is a leaf node with lower magnitude of repeat buyers compared to first-time buyers. From root node one there are three intermediate nodes. The intermediate choice of navigations include consumer who have visited from coupons, retailer-website, and consumers with navigational choices most significantly using shopping-lists. The intermediate category of node six is strongly influenced by beverages, fruits-vegetables and imported-gourmet. Node seven is a leaf node indicating the influence of promotional navigational choice.

Node ten is a leaf node with one intermediate node from the root node. The node is significantly influenced by first-time buyers. It has no significant relationship with any grocery categories, meaning it is independent of grocery items. Node ten is strongly influenced by Bing, Google and other source websites. It is also influenced significantly by submenu/menu (cl), promo, shopping-list and special/seasonal (sp). The left section from root node one which include nodes 2 to 10 have only four choices of navigation, namely, submenu/menu (cl), promo, shopping-list and special/seasonal (sp).

Node 12 is a leaf node and is significantly influenced by repeat buyers. It has one intermediate node from root node. The node 12 is significantly influenced by beverages, household

and imported-gourmet groceries. Node 12 is also significantly influenced by the remaining eight navigational choices, namely, basket, view/review-orders (co), green-basket, Offers, products in basket (pb), product-categories main page (pc), product description (pd) and product-search (ps). Node 13 is significantly influenced by all eight choices of navigations. However, six grocery categories, namely, branded-foods, bread-dairy-eggs, grocery-staples, household, meat and personal-care also have a significant influence. Node 13 has more influenced by repeat buyers compared to first-time buyers. Interestingly, right section has no significant influence by source website which means that only one website has an influence, making the rest of the eight navigational choices redundant.

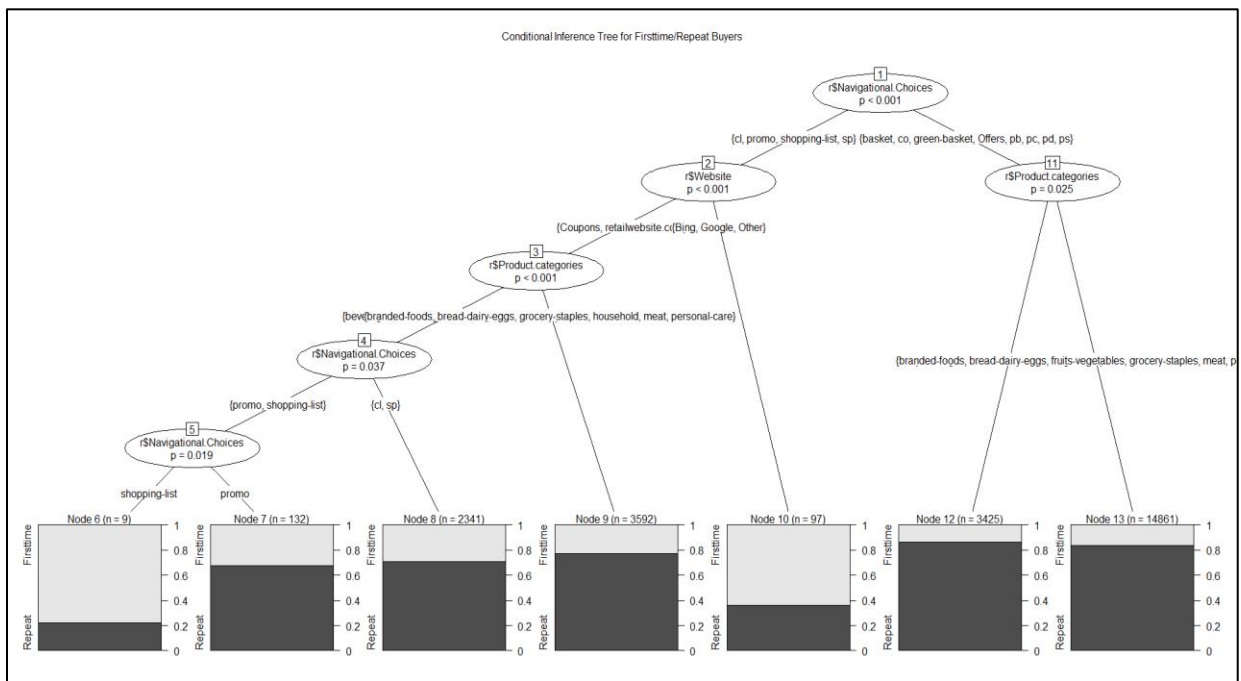


Figure 1: Decision tree for repeat/first-time buyer

Conclusion and Limitations

Basmati-rice is the most prevalent top item. This is expected because data is from the four southern metro cities of India. The existence of cheese within the top item is peculiar as it is not a common item in traditional Indian cuisine. This indicates a lifestyle change influenced by the metro or

international culture. Bread and cheese are the top ten items found in a study conducted by Hulten and Vanyushyn (2011) across France and Sweden. First time consumers exhibit hesitation in using navigational choices, specifically product description and product search. This is indicative of lack of trust in online grocery purchases. Repeat purchases are dependent on trust. Sharif (2014) has found that there is negative moderation of habitual shopping on purchase intention and trust. Apprehensions can lead to non-purchase or non-repeat consumers. The first-time consumers exhibit no clear path but have higher inclination towards browsing of products. This information can be effectively utilized for building a relationship with first-time buyers and offsetting trust issues (Nicolle and Goel, 2013). Retailers can support first time buyers by providing incentives like coupons for next visits or incorporating visitor benefits with the recommendation agent.

Consumer behaviour is a field of study with wide scope. This study has limitations that give future directions for this research. Consumers involved in online purchase behavior have distinct individual personalities. Our data lacked consumer demographics and personality details that could have enabled in creating consumer profiles.

References

Journals

Astudillo, C., Bardeen, M., & Cerpa, N. (2014). Editorial: data mining in electronic commerce-support vs. confidence. *Journal of theoretical and applied electronic commerce research*, 9(1), i-vii.

Baidya, M., & Ghosh, G. (2014). An empirical investigation of repeat buying behavior of consumers of two brands in India. *Journal of Indian Business Research*, 6(3), 255-268.

Bucklin, R. E., Lattin, J. M., Ansari, A., Gupta, S., Bell, D., Coupey, E., & Steckel, J. (2002). Choice and the Internet: From clickstream to research stream. *Marketing Letters*, 13(3), 245-258.

- Chu, J., Arce-Urriza, M., Cebollada-Calvo, J. J., & Chintagunta, P. K. (2010). An empirical analysis of shopping behavior across online and offline channels for grocery products: the moderating effects of household and product characteristics. *Journal of Interactive Marketing*, 24(4), 251-268.
- Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). On the depth and dynamics of online search behavior. *Management Science*, 50(3), 299-308.
- Kolesar, M. B., & Wayne Galbraith, R. (2000). A services-marketing perspective on e-retailing: implications for e-retailers and directions for further research. *Internet Research*, 10(5), 424-438.
- Korgaonkar, P. K., & Wolin, L. D. (1999). A multivariate analysis of web usage. *Journal of advertising research*, 39, 53-68.
- Kuo, Y. F., Hu, T. L., & Yang, S. C. (2013). Effects of inertia and satisfaction in female online shoppers on repeat-purchase intention: The moderating roles of word-of-mouth and alternative attraction. *Managing Service Quality: An International Journal*, 23(3), 168-187.
- Maxwell, J. A., & Loomis, D. M. (2003). Mixed methods design: An alternative approach. *Handbook of mixed methods in social and behavioral research*, 1, 241-272.
- McCarthy, M., O'Sullivan, C., & O'Reilly, S. (1999). Pre-identification of first buyers of a new food product. *British Food Journal*, 101(11), 842-856.
- Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology*, 13(1), 29-39.
- Moe, W. W., & Fader, P. S. (2004). Dynamic conversion behavior at e-commerce sites. *Management Science*, 50(3), 326-335.
- Nicholas, D., Huntington, P., & Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of documentation*, 61(2), 248-280.

- Nicholas, D., Huntington, P., & Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of documentation*, 61(2), 248-280.
- Nicolle, A., & Goel, V. (2013). Differential impact of beliefs on valence and arousal. *Cognition & emotion*, 27(2), 263-272.
- Peterson, R. A., Balasubramanian, S., & Bronnenberg, B. J. (1997). Exploring the implications of the Internet for consumer marketing. *Journal of the Academy of Marketing science*, 25(4), 329-346.
- Shahin Sharifi, S. (2014). Impacts of the trilogy of emotion on future purchase intentions in products of high involvement under the mediating role of brand awareness. *European Business Review*, 26(1), 43-63.
- Sondhi, N. (2014). Assessing the organic potential of urban Indian consumers. *British Food Journal*, 116(12), 1864-1878.
- Van den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research*, 166(2), 557-575.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1-29.
- Wolfenbarger, M., & Gilly, M. C. (2001). Shopping online for freedom, control, and fun. *California Management Review*, 43(2), 34-55.
- Ye, Q., & Fang, B. (2013). Learning from other buyers: The effect of purchase history records in online marketplaces. *Decision Support Systems*, 56, 502-512.
- Zha, X., Li, J. and Yan, Y. (2013). 'Information self-efficacy and information channels'. *Online Information Review*, 37(6), pp. 872–890.

Books

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.