HELPING TARGET GET ON TARGET WITH BIG DATA

Vishwa Bhuta, Miles Sollinger, Bill Su

COMM 4520 – 002 Final Project

March 6, 2015

I. Our Objective

The retail industry has faced challenges from a major disruptive force in the past decade: the Internet. Early adopters such as Amazon have changed the face of traditional retail by harnessing the power of the Internet. Traditional retailers are scrambling to create their own e-tail websites in order to compete. The purpose of our project is to analyze how e-tailers are competing in this new market.

We have chosen Target as our e-tailer for this project because it is one of the most interesting traditional retailers in the United States, known for its adaptability to change. While Walmart eats away a huge share of the discount retail industry, Target has carved a niche market by offering price-conscious customers an aesthetically pleasing experience with well-designed products and superior customer service. Currently, Target is the second largest discount retailer in the country after Walmart, with close to 2,000 stores.

We have two primary objectives. Our first objective is to determine the current state of Target's website and compare the website to its competitors. We want to understand the effectiveness of Target.com in attracting customers to its website and generating revenue for Target. Additionally, Target will be able to understand its position in the e-tailing industry through our competitive analysis. We also want to identify the demographics of customers who shop on Target.com and determine whether the current demographic can generate more revenue for Target.

Based on results obtained from the first objective, our second objective is to provide actionable insights for Target.com to increase revenue by improving the website and targeting attractive customer segments. We hope Target can leverage our analysis to improve its e-tailing competitiveness and achieve revenue growth.

II. Our Method

In our analysis, we chose to compare Target with two of its major competitors. The first one is Walmart, which has invested heavily in its website and achieved success in attracting customers and generating revenue online. The second competitor is Sears, which has been viewed largely as a declining company that has struggled to adapt to both the force of Amazon and Walmart. By comparing Target with one fairly successful and one declining traditional retailer, we can determine Target's success in adapting to the e-tailing market.

The information sources provided are clickstream data from two days in March 2014 as well as demographic data. This small timeframe is a limitation in the data and may be a bias in our results and suggestions. The clickstream data includes the url, duration of page visit, whether the page was a product page, whether the page was a checkout page, whether the page was an exit page, the time of visit, machine id, person id, and session id. The demographic data includes gender, income level, household size, household education, presence of children, and ethnicity. We are going to use this data to create three primary indices to determine the success of Target.com:

1. **Online customer loyalty**: measured by identifying average session visits per each unique customer. (We assume that loyal online customers will make repeat visits to the e-tailer's website.) The customer loyalty of Target and its competitors will be joined with household income level to draw further insight. This analysis will show us both the

overall customer loyalty of those three companies but also the type of customer that is the most loyal.

2. **Checkout churn rate**: determined by the amount of customers ending their session after entering the checkout page. By comparing checkout churn rates across competitors, we can assess if Target needs to make improvements to its checkout page.
3. **Duration per session by gender**: determined by calculating average duration of each unique session visit. We hope to determine customer experiences of Target's website compared to its competitors. We also hope to understand if there is a difference in shopping patterns between men and women on these websites.

III. Are online customers loyal to Target?

Research Question

What is Target's overall online customer loyalty compared to its competitors? What types of customers are most loyal to Target? We want to measure how successfully Target.com attracts repeat customers.

Findings and Takeaways

Overall, Target's customer loyalty is significantly lower than that of Walmart and Sears (**Figure 1**). Despite the fact that Sears has a significantly lower overall session visit count than Target, Sears' customers visit Sears.com more frequently than Target's customers. One explanation for this phenomenon is that because Sears has been targeting low-income customers, it has gathered a group of loyal followers.

A more detailed breakdown of customer loyalty shows that customer loyalty for Target.com is roughly the same across all income levels (**Figure 2**). Comparatively, both Walmart and Sears have a high amount of loyalty from customers in the 15k-25k income group. In particular, Target has significant less loyalty from the 15k-25k income group despite the fact that this group is most loyal to all three websites combined (**Figure 3**). This may be because Target's core customers are in a slightly higher income bracket due to its marginally higher price tag than either Walmart or Sears, which compete on price.

Another interesting finding is that even though people of the 40k-60k income level make the most visits on all websites (**Figure 4**), customers of the 15k-25k income level are the most loyal customers. This shows that customers who make the most visits to the site are not necessarily the most loyal customers. Something for Target to keep in mind is that our data set is biased because the proportion of customers in the 40k-60k income level is statistically higher than expected. (See the Evidence section for more information on this statistical test.) While this does not necessarily disqualify our finding, Target should run these analyses on a more balanced sample in order to confirm our findings' validity.

From the analysis above, we conclude that Target's customer loyalty in general is very low compared to its competitors. At the same time, Target.com does not have a high amount of loyalty in any specific customer income segment. This is inconsistent with Target's overall brand, which has a considerable amount of loyalty from customers. One possible reason for Target's low customer online loyalty may be the instability of its website. For the past few years, Target has experienced multiple website crashes and security breaches during the holiday

season[1]. We recommend Target invest in its network and security infrastructure to provide its customers with a more stable and secure system. In doing so, Target can regain its customers' loyalty and trust.

After stabilizing its website, we suggest Target focus on raising online customer loyalty by offering more online-only deals. Specifically, we recommend targeting the 40k-60k income group because this group has a high number of website visits and reflects Target's core customer. Even though customers of the 15k-25k income level are most loyal to online e-tailers in general, this segment is not Target's core customer.

Evidence

We ran a chi-square goodness of fit test on Target's number of customers per income level and found that the observed number of customers per income level differs significantly from the expected proportion ($p < 0.01$). Specifically, the proportion of customers in the 40k-60k income level was statistically higher than expected, representing 25% of the chi-square statistic.

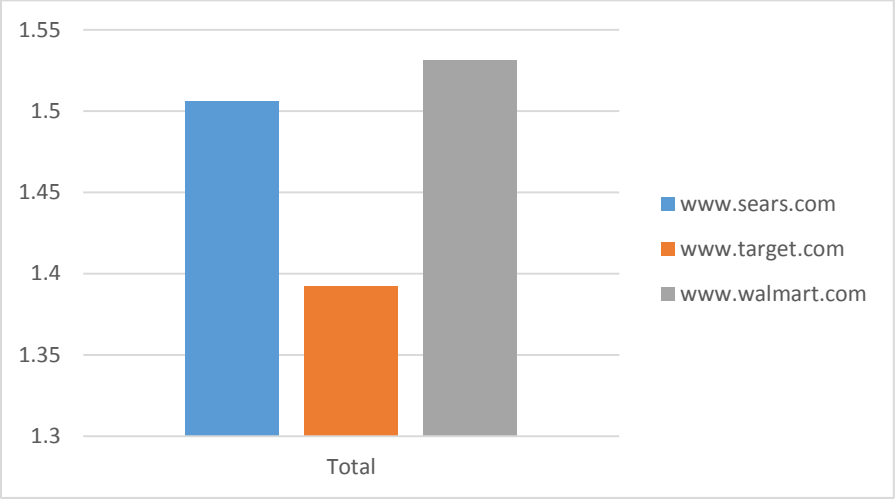

Figure 1. Overall Customer Loyalty

[1] http://www.nytimes.com/2014/03/06/business/a-top-target-executive-resigns.html?_r=0

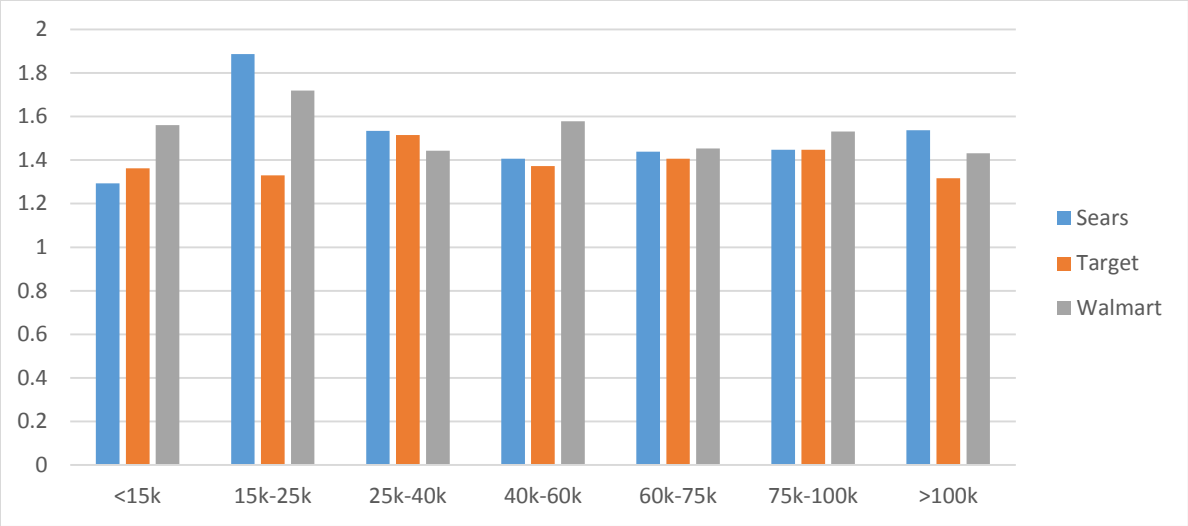

Figure 2. Customer Loyalty by Income Level and Website

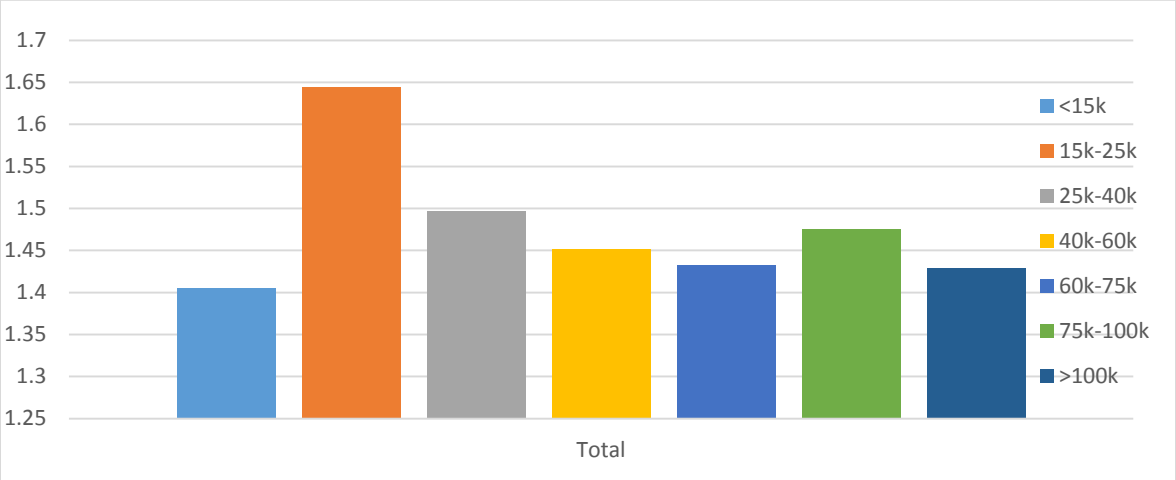

Figure 3. Customer Loyalty by Income Level

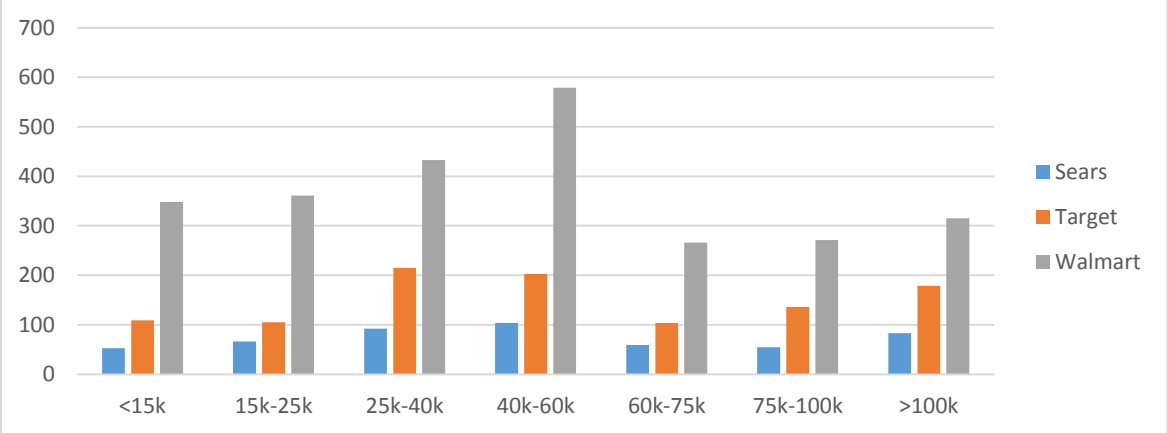

Figure 4. Total Sessions by Income Level and Website

Technical Description

In order to obtain the count of sessions each unique person has visited, we first grouped the clickstream Bigsheet by personid, sessionid, and urlhost. Then we grouped the result by urlhost and personid to obtain the count of sessions. To include demographic information, we conducted an inner join between the filtered clickstream sheet and demographic sheet based on the common variable personid. Finally, we grouped the variables by urlhost, hhincomeid and obtained count for personid and sum for sessionid.

IV. What is Target's Checkout Bounce Rate?

Research Question

The research question for our second insight is: what is Target's checkout bounce rate and how does the rate compare to competitors? We wanted to investigate this question to understand the performance of Target's checkout page. By comparing the checkout bounce rate across competitors, we can assess if Target needs to make improvements to its checkout page.

For this analysis, we did not include Sears as there are not enough observations for our test to be statistically significant. We included Macys instead of Sears, as Macys is also a retailer with a significant online presence.

Findings and Takeaways

We found that Target's checkout bounce rate is 12.07% and Target is in a moderate position compared to its competitors. Walmart leads the pack at 5.14%, while Macys lags behind at 23.08%. We also found that the duration of the visit on the checkout page is highly positively correlated with the checkout bounce rate ($r = 0.99$). **Table 1** summarizes these findings.

The main takeaway is Target needs to improve its checkout page if it hopes to compete with Walmart in the e-commerce space. Target had a significantly higher number of checkout page exits than Walmart. The correlation between checkout bounce rate and page visit duration suggests that confusion with the checkout page, possibly stemming from issues with the checkout process or functionality of the checkout page, may be causing users to exit their session. If Target does choose to improve its checkout page, it should do so through A/B testing to measure the effectiveness of any proposed site changes. A/B testing would minimize the risk of altering the checkout process, which is an extremely important process for e-tailers.

However, if Target is not concerned with competing with Walmart in e-commerce it may choose to do nothing. Target's checkout bounce rate is significantly better than Macys' (see the Evidence section for the statistical test output). Additionally, the Baymard Institute, a leader in e-commerce usability research, estimates the average e-commerce cart abandonment rate to be about 68%.[2] While the checkout bounce rate is not exactly equal to cart abandonment rate, one can be used as a proxy for the other due to the absence of transaction data. With this in mind, Target is well below the industry average and is not necessarily in a bad position.

[2] http://baymard.com/lists/cart-abandonment-rate

Evidence

Company	Checkout Bounce Rate	Avg. Duration on Checkout Page
Walmart	5.14%	27.43
Target	12.07%	35.04
Macys	23.08%	53.85

Table 1. Checkout Bounce Rate and Avg. Duration on Checkout Page Across E-Tailers

To assess the validity of this finding, we ran a univariate analysis of variance (ANOVA) of number of exits across e-tailer competitors. The overall difference in exits is significant with $p < 0.01$. The post-hoc analysis, which runs a t-test between all pairs of e-tailers, shows the p-value is < 0.01 for all pairs of e-tailers. Therefore we can conclude there is a significant difference in number of exits across all three e-tailers. See the Appendix for the output of this statistical test.

Technical Description

To investigate this question, we looked at instances in the clickstream data in which the user was viewing a checkout page. We filtered out only the relevant websites then filtered so we were only looking at checkout page views. The checkout bounce rate for a given e-tailer was calculated as the number of session exits divided by the number of checkout page views.

V. How long do people spend on Target.com per session?

Research Question

The last research question we wanted to explore was: how much time do people spend on Target's website per session, and does this duration vary by gender? How does this duration compare to competitors?

Findings and Takeaways

As can be seen in **Figure 5**, both men and women are spending more time on Target's site per session than on Walmart or Sears' site. Interestingly, men appear to spend longer on Target's website than women; the opposite is true for both the competitors. Running an independent t-test on gender, we see that across the three competitors, there is not a significant difference in duration by gender. This means that despite what we may infer on first glance, men and women are spending similar amounts of time on each e-tailer's website.

There are two explanations of the results. One explanation is that people are using Target's website for browsing purposes, which is keeping them on Target's website longer per session. Customers may go onto the Sears' or Walmart website with an idea of what they want, therefore spending less time browsing. If this is true, Target should focus on how it can convert the browsing behavior into sales. This might be done by creating better calls to action or more attractive product pages that might tempt customers to make a purchase. An alternative explanation for customers' behavior is that Target's website is harder to use than its competitors'. This would mean that customers are staying on the website longer because they

cannot find what they are searching for. In this case, Target should benchmark its competitors' website to see what makes their site more efficient.

Without further analyses, however, we cannot determine which (if either) of these explanations is true. In future analyses, we could look at how many pages people are looking at per session and then connect that to duration per session. If people are looking at a few pages for a long time, it might be because the pages are too complex. However, if people are looking at many pages for a long time, this could indicate browsing behavior. We can also consider integrating transaction data into this analysis to understand if duration per session is related to conversion rates. If people are staying on the website for a long time, are they actually making a purchase at the end? If they are, does it matter how long they spend on the website before they buy something? If they aren't, what is Target not doing that competitors may be doing to end a session with a purchase? Adding transaction data into the mix would add significant value to the quality of analyses we could conduct.

Evidence

Figure 5. Duration per session by gender

We ran a t-test of duration between genders and found there is no statistically significant difference in duration. See the Appendix for the statistical output of this test.

Technical Description

To investigate this question, we aggregated the duration per page by session id to see how long users were spending on the site per session. We filtered out only the relevant websites then calculated the total duration per session using a group by function; we also grouped by gender to see how duration per session differed by gender. To see the average duration per session for each gender for each site, we calculated an average duration variable using a group by function. Finally we used a pivot table to visualize the data.

VI. Conclusion: Targeting A Brighter Future

Based on our analysis, we have concluded that Target's e-tail division is not optimized for attracting customers. First, Target's customer loyalty is significantly lower than its competitors due to constant website crashes and security breaches. Second, although Target's checkout bounce rate is lower than some of its major competitors and is below the industry average, Walmart is still ahead. Finally, we found that customers spend on average more time per session on Target.com compared to its competitors. Although the reason for this is unclear, there is a possibility that it is due to a user unfriendly site.

We have four final recommendations to improve Target's position in the e-tailing industry and further these analyses:

1. **Offer specialized online deals to build customer loyalty**: We recommend Target offer online-only promotions and discounts to build customer loyalty in the online space. Specifically, Target should focus on the 40k-60k segment because these customers account for the most visits on Target's website and this segment fits within Target's core customer. Target should also run these analyses with a balanced sample to reduce any bias from having more customers in any particular income segment.
2. **Improve customer experience through infrastructure development:** We recommend Target invest in its web infrastructure to offer a more stable and secure online shopping experience. Target needs to reverse its trend of website security issues. This is essential for retaining customer trust for the website.
3. **Improve checkout page design and functionality to reduce checkout bounce rate:** If Target hopes to compete with Walmart in e-commerce, it will need to lower its checkout page bounce rate. The higher average page duration on the checkout page suggests there may be problems with the checkout page's design or functionality and customers are not able to successfully checkout. Target should address this issue to reduce its checkout bounce rate and increase online revenues.
4. **Integrate transaction data and more data into the analysis**: Currently our data only allows us a glimpse into the first level of analyses regarding Target's website. We need to understand how the website activity relates to sales conversions in order to analyze the data at a deeper level and make specific, empirical suggestions. Furthermore, the narrow timeframe of the data skews any seasonal trends in the data, so running these analyses with data that stretched over a longer time frame (eg. a year) may be crucial in validating our suggestions.

VII. Appendix

Big Sheets:

Insight 1: ClickStreamPreJoin.csv, ClickStreamJoinedWithDemo.csv

Insight 2: No Big Sheet, analysis was done in Excel

Insight 3: DurationPerSessionRawData.csv

Spreadsheets and Pivot Tables:

Loyalty.xls: Contains customer loyalty, customer count and sum of sessions sorted by income level and urlhost. There are four primary Pivot Charts within this Excel document: Overall customer loyalty sorted by urlhost, customer loyalty sorted by urlhost and income level, customer loyalty sorted by income level, and total session count sorted by income level. Also contains the chi-square test.

Checkout Bounce Analysis.xlsx: Contains the checkout bounce rate calculations for Insight 2 as well as the ANOVA statistical test.

Duration By Gender.xlsx: Contains the duration per session analysis as well as the t-tests.

Statistical Output:

ANOVA Output

This ANOVA compares number of checkout page exits across e-tailers.

One factor ANOVA

Mean	*n*	*Std. Dev*	
0.1	954	0.22	Walmart
0.1	290	0.33	Target
0.2	52	0.43	Macys
0.1	1296	0.26	Total

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Treatment	2.40	2	1.200	17.93	2.08E-08
Error	86.49	1293	0.067		
Total	88.89	1295			

Post hoc analysis

p-values for pairwise t-tests

		Walmart 0.1	Target 0.1	Macys 0.2
Walmart	0.1			
Target	0.1	.0001		
Macys	0.2	1.25E-06	.0048	

T-test Output

This t-test compares the total duration of Sears, Target, and Walmart across genders.

Hypothesis Test: Independent Groups (t-test, pooled variance)

Male	Female	
948.682610	1,018.861358	mean
249.304493	105.566846	std. dev.
3	3	n

4	df
-70.1787475	difference (Male - Female)
36,648.5445402	pooled variance
191.4380958	pooled std. dev.
156.3085507	standard error of difference
0	hypothesized difference
-0.449	t
.6767	p-value (two-tailed)

Chi-Square Goodness of Fit Output

This chi-square test compares the observed number of Target customers per income level to the expected level. The expected level assumes an equal proportion of customers per income level.

Goodness of Fit Test

observed	expected	O - E	$(O - E)^2 / E$	% of chisq
80	107.571	-27.571	7.067	11.67
79	107.571	-28.571	7.589	12.53
142	107.571	34.429	11.019	18.19
148	107.571	40.429	15.194	25.09
74	107.571	-33.571	10.477	17.30
94	107.571	-13.571	1.712	2.83
136	107.571	28.429	7.513	12.40
753	753.000	0.000	60.571	100.00

60.57	chi-square
6	df
3.45E-11	p-value