Harris Teeter
Neighborhood Food & Pharmacy

MARKETING RECOMMENDATIONS

Gang Chris
Hannah Emme
Paramita Ghosh
Mykayla Goodwin
Ali Hasan
Anna Korosteleva

Table of Contents

Introduction

Industry Overview

Grocery Retailing

Grocery sales are projected to exceed $700 billion by 2020, a 14% increase from 2015, which is valued at $635.8 billion (Mintel, 2016). The US grocery retailing industry is highly concentrated, with the top 50 companies generating more than 70% of total revenue (Hoovers, N.D.). Profitability is heavily reliant on high volume sales, and profit margins are being further reduced due to increased competition with non-traditional retailers. Typically, these retailers are able to offer consumers price points below traditional grocery retailers due to their large size and infrastructure, but other indirect competition also comes in the form of subscription-based grocery services such as Amazon Fresh and fresh ingredient delivery services such as Blue Apron and Hello Fresh. In the graphic below, MULO (multi-outlet) retailers include mass merchandisers such as Target, warehouse clubs such as Costco, and even dollar stores.
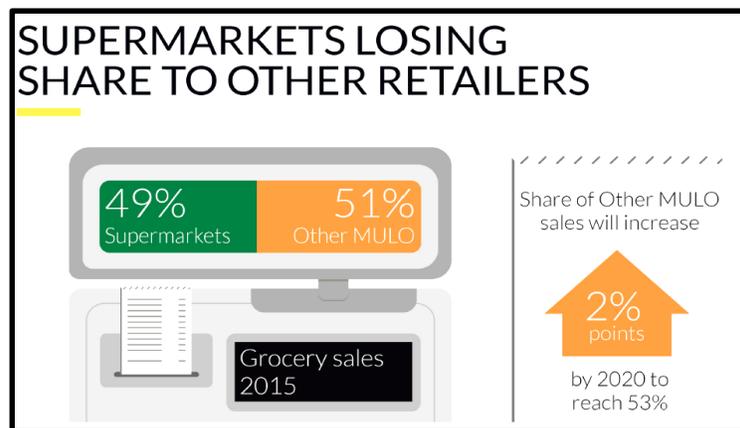
Grocery Retailing Infographic Overview, Mintel Group, 2015

Online Shopping Meets Grocery

Though traditional brick-and-mortar stores remain the primary channel for retailers, online and mobile commerce are increasing in usage, now representing a significant portion of retail sales. Even for consumers that prefer online shopping, physical storefronts serve as "showrooms" so that customers can

interact with products prior to purchasing. Similarly, those that prefer to go in-store to make purchases often research products online before checkout (PwC, 2016).

The grocery industry has been changing rapidly over the past decade and online grocery shopping sales are expected to grow nearly 10% annually for the next three years (Mintel, 2016). This seems to be an indication that grocery retailers are looking for other avenues to expand their market share.

Grocery Competition in the Charlotte Market

Similar to overall industry trends, Charlotte is marked by an extremely competitive grocery retailing landscape. Publix entered the North Carolina market in 2014 and plans to expand even further throughout the state in the coming years - they have already opened 14 stores in the Charlotte metropolitan area (Peralta, 2016). According to Sarah Peralta, a contributor for the Charlotte Observer, many other companies also plan to enter or expand operations in Charlotte, such as Germany's Lidl, a grocery store similar to Aldi's, which plans to open its first store in the area in 2018.

The Charlotte grocery market is relatively saturated; new entrants and competitor expansion has only further diluted market shares for all players in the industry. This means that consumer dollars dedicated to grocery items are becoming more and more fragmented as they are now spent in a variety of places.

Harris Teeter Overview

Harris Teeter is headquartered in Matthews, North Carolina and became a subsidiary of Kroger, the nation's largest supermarket operator, in January 2014 (Peralta, 2015). The grocery retailer operates through 235 stores across 9 states in the southeast region and recently reclaimed the top grocery market share in the Charlotte market, holding 19.4 % of the market (Peralta, 2016). Though market share declined slightly, Walmart remains the largest grocery retailer in the region due to multiple storefront options - Sam's Club, Neighborhood Market, and Supercenters.

Charlotte-Area Grocers By Market Share		
Store	**2015 Market Share**	**2014 Market Share**
1. Harris Teeter	19.4%	19.7%
2. Food Lion	19.1%	19.2%
3. Walmart Supercenter	19.0%	19.8%
4. Publix	6.0%	4.8%
5. (Tie) Sam's Club	5.6%	6.0%
5. (Tie) Bi-Lo	5.6%	5.8%
7. Costco	3.1%	3.3%
8. (Tie) Super Target	2.4%	2.4%
8. (Tie) Aldi	2.4%	2.6%
10. Neighborhood Market	2%	1.4%

Martin, 2016. http://www.bizjournals.com/charlotte/news/2016/04/27/hometown-grocer-notches-win-in-market-share-battle.html

The team reached out to Harris Teeter consumer researcher Ken Hayden to see if there was any analysis we could provide that would be of benefit to him or if there was any information he could provide to assist with the research. Hayden commented that Harris Teeter began offering consumers online shopping options approximately 15 years ago, though it has only started gaining traction in the last five years. Today, Express Lane is available in 75% of Harris Teeter locations (Hayden, 2016), primarily in large urban cities such as Washington D.C. The core customer base is composed primarily of shoppers with kids, customers that are unable to leave the home due to being disabled or elderly, and those without personal transportation to carry large quantities of groceries.

Hayden disclosed that the Express Lane customer interface is managed by a third party, so data specific related to the program is unavailable, though Harris Teeter is able to access data about what items customers are purchasing. He also mentioned that the company has received an unspecified number of

complaints in regards to the following topics:

- Payment functionality within the app – Customers are currently unable to pay for their purchase through the app, requiring them to go in-store when picking up orders.

- Waiting in line – Associates are not always able to assist Express Lane customers immediately upon arrival, so despite using the service, customers are still required to wait in line in order to checkout.

- Quality of selected items – Because employees make the selection for customers, occasionally there is a difference in opinion about the quality of fresh foods. Additionally, weight differentials between the weight selected for purchase online and the actual weight of the item slightly impacts the total amount due at checkout. For instance, if the customer orders a pound of grapes online, the bunch received in their order may be slightly more or slightly less than one pound, changing the price quoted online.

- Weekly ad aspects - The aspects for advertisements listed on the website are different from the app, where the zoom feature is not available, so customers have to click through several pages to view the complete advertisement.

Hayden also mentioned that customers like the in-store experience to see new products available in stores and to make price and nutritional label comparisons, all of which cannot be done through the app. Next, we reached out to Chase Cabanillas, the director of digital technology and operations, who informed the team that the basket size for customers at Harris Teeter is approximately three times higher online than in-store, with the average receipt in-store totaling less than $50 compared to more than $100 online (Cabanillas, 2016).

Research Problems & Objectives

With the increasing role online retailing plays in consumer shopping, it is important for Harris Teeter to prioritize their online presence in order to remain competitive in the Charlotte market. An interview with Harris Teeter consumer researcher Ken Hayden provided a preliminary foundation upon which the team developed our main objectives. As previously discussed, Hayden disclosed that because

the Express Lane customer interface is managed by a third party, data availability is limited and would need to be collected by the team.

Hayden's feedback solidified the team's interest in exploring the online retailing portion of the grocery industry and Harris Teeter specifically, due to the company's strong local presence. Our research objectives included collecting consumer opinions about the Express Lane feature, both positive and negative, and identifying specific customer demographic factors that might influence participation in the online program. Our goal is to identify what consumers consider to be the problems associated with online grocery shopping and make recommendations for Harris Teeter to not only increase participation in the Express Lane service, but to improve the overall shopping experience across all of their channels.

Analysis Methods & Results

The quantitative data used in the research was generated through use of an original survey. Though internal data provided by Harris Teeter was desired, we ultimately had to rely on primary data for regression purposes. Qualitative data consisted of text-based responses collected in the survey, publications about grocery retailers in Charlotte, as well as consumer reviews or posts on Yelp.com and Facebook. Web scraping allowed the team to utilize numerous analysis methods in order to develop more well-rounded and actionable insights.

Online Survey

To begin the analysis process, our team conducted a survey through Qualtrics in order to help identify improvement opportunities for Harris Teeter's online shopping platform. The survey was distributed primarily through social media and email to a convenience sample of contacts at the national level. Respondents were primarily concentrated in the Charlotte area, aligning with the local focus of the research objectives. The pool was almost evenly split in gender, with 76% of participants falling between the ages of 21-39.

The survey began by capturing basic demographic information before moving into questions about online shopping in general. These were designed to gauge shopping behaviors and preferences, as well as to identify factors that would influence shoppers to choose one online retailer over another. The

team found that participants were most likely to shop for *electronics* and *clothing/footwear* online (68%) and that *free shipping* (71%) or *lower prices* (69%) would entice them to choose a different retailer.

Next we focused on the online grocery retailing industry, where we asked questions about their experience with grocery shopping online, changes in behavior over the last year, as well as questions to help us identify why they buy groceries online. The survey results reported low participation rates in online grocery shopping, with only 11% indicating that they currently shop for groceries online. As shown in the graph below, *grocery* was selected as the #1 category they are least likely to shop. Those that did choose to shop online said they did so because of convenience and the time saving benefit.

Retail Category Least Likely to Shop for Online

- Clothing & footwear (12.05%)
- Grocery (34.94%)
- Jewelry (15.66%)
- Homeware & furniture (4.82%)
- Appliances (14.46%)
- All Others (18.07%)
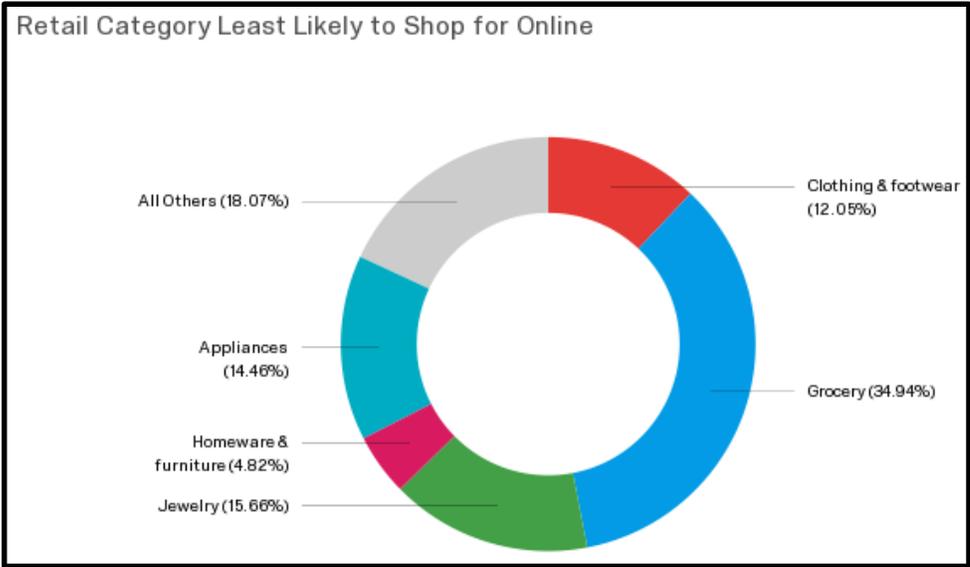
When asked why they did not shop online for groceries, 21% of participants indicated that they preferred to select their own fruits and vegetables and appreciated the ability to see the product prior to purchase. This is consistent with our survey results that *availability of fresh produce* is the most important attribute for online grocery shopping and information shared during the interview with Hayden.

In general, the brick-and-mortar stores remain the channel of choice in terms of retail touch point, but results did also indicate that online grocery shopping has at least made people curious enough to try it. Of those surveyed, 10% have bought groceries online in the last 30 days and 27% did so in the last 12 months. In fact, 13% said they actually shopped for groceries online more often than they used to a year ago, even if that was not their overall shopping preference.

Regression

The next step in analyzing the survey data was to use binomial logistic regression to discern any significant independent variables in determining whether a shopper preferred to shop in-store or online. Because only one survey participant indicated that they grocery shopped online and another ten indicated their shopping is characterized by a mix of both online and in-store, the team combined the two groups due to their shared online activity. As a result, *online* represents both categories.

After running a series of regressions (Appendix A1-4), we discovered that there were a few factors within the survey data that were either significant or close to an acceptable significance level (between 0.1 and 0.2) across each different regression regardless of the other independent variables. When combined, these variables became even more significant. It is important to note that due to our dependent variable, in which 1=in-store and 2=online, the regression measured the probability that a respondent prefer to shop in-store. Therefore, we have to carefully interpret data to apply to online preferences. In this regression output, utilizing a significance cut-off level of 0.1, the following independent variables are significant:

- Household_ : Number of people in the household. This variable has a positive parameter estimate, indicating that a larger number of household members increases the likelihood that a customer prefers to shop in-store. This seems counterintuitive based on our other analysis results, however, the survey sampling bias related to the majority of respondents reporting household sizes of 2 and 3 people , suggests that this data may not be applicable to a family with multiple children.

- Shop_Toys : Respondent indicated that they shopped for toys online. In this case, a negative parameter indicates that those who already shop for toys online are less likely to prefer shopping in store.

- WhyGrocOnline_AvoidC : Respondent indicated that they shopped for groceries online to avoid crowds at the store. This parameter also had a negative sign, reinforcing the idea that someone who shops online for groceries to avoid crowds is less likely to prefer shopping in-store.

However, it is more important overall that it is statistically significant as a reason for shopping online, while other reasons were not.

Parameter Estimates					
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	1.8724	1.6649	Infty	1.12	0.2607
Household_	0.8916	0.5228	Infty	1.71	0.0881
Shop_Toys	-2.2135	0.9911	Infty	-2.23	0.0255
RetailChoice_BrandLo	1.6428	1.1102	Infty	1.48	0.1389
RetailChoice_FreeShi	-1.6095	1.3984	Infty	-1.15	0.2497
GroceryTripsPerMonth	0.1547	0.5861	Infty	0.26	0.7918
WhyGrocOnline_Budget	0.9832	1.5853	Infty	0.62	0.5351
WhyGrocOnline_AvoidC	-2.9916	1.0732	Infty	-2.79	0.0053

Yelp Text Clustering

To begin exploratory analysis, we used R to scrape data from Yelp.com specific to the 47 Harris Teeter stores in the Charlotte metropolis. The team investigated cluster sizes ranging from three to ten and descriptive terms of 15 or 20. Text clustering was performed without providing a start or stop list. After reviewing the output of numerous combinations, the team selected four as the optimal number of clusters based on the low RMSSTD, the pooled standard deviations of all of the variables that form the cluster, with 15 descriptive terms. This combination also produced more specific clusters, making them easier to label. Results from each clustering attempt can be found in the appendix.

Yelp.com Harris Teeter Reviews

Cluster	Descriptive Terms	Frequency	Percentage	RMSSTD
1	+'salad bar' +bar +salad +park +'park lot' starbucks +lot hot +fresh +area +selection +meat +wine always +great	138	25%	0.13020202
2	+nice +small +neighborhood +shop ht +clean +big +item +wine +location +find +store pretty +selection +good	157	28%	0.12779192
3	'Customer service' +customer +service +offer back +want +employee +day +know +time harris +shop +look +friendly deli	183	33%	0.13752342
4	Quality +lack +bag charlotte +high +price food +great staff +'grocery store' deli hot +know +grocery +food	84	15%	0.12840967

As depicted in the four cluster summary above, the cluster with the greatest frequency is cluster three, representing 33% of the reviews. This group mentions *customer service* twice, including terms such

as *employee* and *friendly*, possibly suggesting that Harris Teeter has a very high level of customer service across their Charlotte stores. The second cluster is the next largest group and contains reviews that seem to focus on the visual aesthetic of the stores, listing the store as pretty among other positive descriptors. The first cluster emphasized the salad bar and parking lot. Unlike the first three groupings which were easier to label, the fourth cluster produced slightly more ambiguous terms and had the lowest frequency, representing 15% of the Yelp reviews.

This iterative approach to the unstructured analysis provided the team with insights about aspects of the grocery retailing business that Harris Teeter does well, rather than what the company could improve upon. First, the clustering revealed that consumers have a positive perception of the visual aesthetic of the stores, commenting that they are "pretty" and "nice." Second, Harris Teeter appears to excel in the general cleanliness and upkeep of the stores. Regardless of cluster size, all of the analysis produced similar results in that there was a cluster with high frequency referencing the customer service, one referencing the fresh food offerings, and another referencing the size and cleanliness of the stores. This helps solidify our hypothesis that these are areas that Harris Teeter excels in and acts as a differentiator in the industry.

Yelp Sentiment Analysis

Sentiment analysis is used to figure out the impressions about Harris Teeter from customer. This analysis is powerful because it saves times that using PC to identify 500 tones instead of human reading.

After conducting the sentiment analysis by running self-constructed R code, there are 427 positive, 45 neutral, and 87 negative comments. Therefore, customers who left comments on Yelp satisfied with Harris Teeter.

Yelp Regression Models

Using regression helps to better understand consumer sentiment. Part of our sentiment analysis was to conduct linear regression on the data collected from Yelp. The rating that customers assigned to their experience with Harris Teeter customer service and store visit was of prime importance as this rating was selected as the dependent variable.

Topic	Terms	Parameter labeled in regression
1	+bar, +salad, +salad bar, +pizza, +lunch	TextTopic_raw1
2	+grocer, +checkout person, person, +market, +checkout	TextTopic_raw2
3	+customer, customer service, +employee, +cashier, +manager	TextTopic_raw3
4	+meat, +seafood, +bread, +cheese, +fruit	TextTopic_raw4
5	Staff, friendly staff,+cashier,stuff, +pizza	TextTopic_raw5
6	+wine, +beer, +park lot, +food, +employee	TextTopic_raw6

Other than these topic variables, there are some numeric variables, such as how many funny ratings, cool ratings, and useful ratings are given from other Yelp user. The table below shows the linear regression results. The dependent variable is the overall rating from the customer.

Yelp.com Sentiment Analysis - Linear Regression

Model Fit Statistics					
R-Square	0.1226	**Adj. R-Square**	0.1147		
Parameter	**DF**	**Estimate**	**S.E.**	**T-Value**	**Pr> |t|**
Intercept	1	3.3886	0.0757	44.78	<.0001
TextTopic_raw1	1	0.9653	0.3698	2.61	0.0094
TextTopic_raw2	1	-2.0225	0.3233	-6.26	<.0001
TextTopic_row5	1	1.1963	0.4869	2.46	0.0144
useful	1	0.1013	0.0363	2.79	0.0055

We also developed a second logistic regression model to better understand the data through sentiment. After clustering, the data was partitioned into two models - one with a positive sentiment and the other with a negative sentiment. To prepare the model for regression, we utilized dummy coding to reflect either a positive or negative sentiment and removed values such as customer names, store ID numbers, and addresses. Variables such as price, reviews, ratings and votes were used for the review itself, indicating readers' opinions regarding whether the review was useful, funny and/or helpful.

Positive Sentiment Model

Parameter	DF	Estimate	S.E.	Pr> |t|
Intercept	1	2.3905	0.6705	0.0004
TextTopic_raw2	1	1.2546	0.6684	0.0605
TextTopic_raw3	1	1.7814	0.8747	0.0417
TextTopic_row5	1	1.9395	1.0720	0.0704
price	1	-1.7335	0.2944	< 0.0001
rating	1	0.4847	0.0953	< 0.0001

Negative Sentiment Model

Parameter	DF	Estimate	S.E.	Pr> |t|
Intercept	1	2.0216	0.8532	0.0178
TextTopic_raw1	1	2.2002	0.8608	0.0106
TextTopic_raw2	1	-1.9075	1.1020	0.0835
TextTopic_raw5	1	-3.9032	2.0525	0.0572
TextTopic_row6	1	-3.1667	1.7980	0.0782
price	1	-0.7467	0.3615	0.0389
rating	1	-0.6263	0.1175	< 0.0001

Our findings reveal that the ratings from the customer reflected the overall sentiment of their comment. In both of the models, the rating is significant as the p-value is small. A high rating score from a customer indicated a high probability that the comment will be positively stated. Conversely, a low rating score from a customer indicated a high probability that the comment will be negatively stated. Positive Terms included those such as *customer service, staff* and *services.* Examples of negative terms include *salad bar*, *salad* and other foods, reflecting lower satisfaction from a customer's viewpoint.

Facebook Text Clustering

The team collected 1,583 Facebook posts from the Harris Teeter company page from January 1, 2015 through April 12, 2016 for analysis. Posts included those made by Harris Teeter, page followers or

visitors, and comments made by page visitors on individual posts. The main goal of this data analysis was to determine what people like about Harris Teeter, what they do not like, and what people are saying about online shopping and Express Lane on Facebook.

Facebook data was separated based on the number of *likes* because of the features ability to enable page viewers to show instant reactions and emotion when viewing a post. A post that is *liked* by a large number of people is more engaging and therefore has a further reach than a post that is *liked* by less people. Additionally, when compared to the *share* and *comment* features, *like* is more engaging because it takes less time and effort for visitors to do.

The team divided the data into five groups, treating each as a separate dataset for text clustering. We made ten clusters from each of the five separated datasets, resulting in a total of 50 labels. After labeling them, the team compared the cluster labels among each group and developed the following insights:

- The label *coupon and sales* appears in nearly all of the groups, while *fresh food* and *fresh local produce* are present in all of them.
- Highly liked groups contain the labels such as *online shopping by moms*, *Express Lane with fresh produce*, and *good employee behavior,* as well as mentions of organic food.
- *Sweepstake* appears only in the least liked group.

The Facebook clustering provided insight as to what customers like and do not like about Harris Teeter. Interestingly, sweepstakes appear to not be very popular among Harris Teeter customers, suggesting that this is not an area they should focus on as a differentiator from competitors. Freshness of food, organic options, and coupons appear to be well liked as expected.

Express Lane Reviews

To better understand what customers think about Harris Teeter's Express Lane program, the team identified 107 reviews that specifically discussed the online shopping feature from Facebook and Yelp. The data was gathered through web crawling, and was then merged and analyzed together.

Complete process flow of Express Lane Reviews
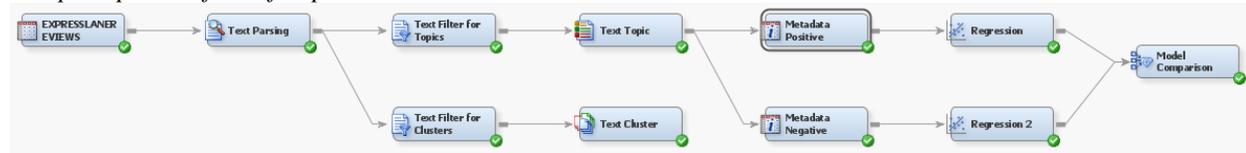
       After the initial text-filtering run, we found numerous irrelevant terms scattered throughout the results, such as dates listed in Harris Teeter advertisements and URLs. We manually removed these to clean the data and improve the analysis. Terms removed included those such as *http://bit.ly/vp75po, 8th, bit,* and *X80.* The cleansing improved the results in terms of meaning and relevance.

Express Lane Text Topics & Text Clusters

       In order to determine the optimal number of topic categories, the team attempted multi-term topics ranging from five to ten. Comparison across the results indicated that fewer topics generated the best results. Additionally, removing irrelevant terms improved results tremendously, producing the final model as displayed below.

Topic ID	Doc. Cutoff	Term Cutoff	Topic	# Terms	# Docs
1	0.216	0.108	+item, +special, +coupon, weekly, +love	22	17
2	0.285	0.102	September, +card, vic, gift card, july	18	5
3	0.240	0.104	+platter, party, +party platter, +order, +holiday	16	13
4	0.252	0.107	Lane, express, +grocery, +shop, +start	23	17
5	0.173	0.106	Harris, teeter, +store, +offer, +location	21	10

      Based on the keywords in each of the five topics, the categories can be described each in the following ways:

              Topic 1 - Specials & Coupons
              Topic 2 - Monthly Promotions
              Topic 3 - Holidays & Parties
              Topic 4 - First Express Lane Experience
              Topic 5 - Store Options & Location

       Next, we clustered Express Lane specific data, attempting cluster numbers from four to ten, each with ten descriptive terms. Comparison across results again indicated that five clusters provided more useful grouping. The results mirrored those of the text topic analysis and can be seen in the table below.

Cluster	Descriptive Terms	Frequency	Percentage
1	+'party platter' +platter football http://bit.ly/vp75po party +holiday +order local 'order online' ready	18	17%
2	Shop fresh +location +area +offer +bag +time +card +shopper service	15	15%
3	+manager +item +coupon +customer nc +want +find +store ht service	30	29%
4	+'weekly special' love +'online shop' +work great +happen +special +shop +'grocery store' +love	17	17%
5	1kp3jfw bit ly shopping express lane nhttp x99ll xe2 +week	23	22%

Because the team gathered similar results from the text topics and text clusters, we selected the text topic analysis due to its more inclusive results. With clusters, each document is assigned to only one category. Text topics, on the other hand, assigns each document to zero or more categories. Because reviews vary in length and can range from one sentence to several paragraphs, it makes sense that each review could potentially belong to several categories. From the analysis, the team identified five major categories that Express Lane customers feel strongly about, which Harris Teeter could consider for improvements to the program:

- Free or discounted shipping
- First-time purchase incentives
- Promotions for holidays, parties, games
- Customer service
- Physical store location
- Technical component of Express Lane

Express Lane Sentiment Analysis

In addition to clustering and text topic analysis, the team wanted to derive insight from the sentiment of Express Lane reviews on Facebook using R. The team discovered that 1,030 of the reviews were positive, 213 were negative, and 340 were neutral. Overall, the Facebook sentiment is highly positive, with roughly only 20% of the sentiment identified as negative. This suggests that customers are generally happy with Harris Teeter. In the last two years, approximately 1,500 people took the time to comment, complain, praise, and share their Harris Teeter experience.

To analyze the sentiment, we first worked with text topics and text clusters for all of the comments. Because there was not a large amount of text data, we ran clusters and text topics with 3 to 20 clusters/topics each, as well as various settings for SVD resolution and algorithms. The best results were those with fewer clusters, at which point patterns for specific topics began to emerge. The best model contains four text clusters and was run with low SVD resolution and a hierarchical clustering algorithm shown below.

Cluster	Descriptive Terms	Conclusions
1. Express Lane & Customer Service (31%)	Purchase, express, Harris, lane, Teeter, service, sale, national, customer, shopping, day, today, week, HT	In addition to various positive comments about Express Lane, Harris Teeter's customers enjoy the customer service and shopping at the stores.
2. Location-Based Promotions (24%)	Super doubles, coupon, coupons, super, VA, doubles, deals, items, store, area, stores, fresh, people, NC, free	"Super doubles" is a hot topic, because occasionally Harris Teeter runs double coupons for a few days, but not at all stores. A lot of promotions are specific to certain stores or states. Customers care about location because those who don't get the promotions for their local store get upset.
3. Freebies (23%)	Month's taste, cake, card, comment, enter, gift, month, prize, win, chance, details, click, share, sweepstakes	While the term "month's taste" gets picked up in this cluster, "Taste & Tell" is what customers care about. Through Taste & Tell, reward cardholders have an opportunity to try new products each month for free. People like to save money, as well as free items and gifts.
4. Harris Teeter Promotions (22%)	Dinner, wine, recipe, easy, perfect, cheese, favorite, always, taste, recipes, great, day, service, time, delicious	Harris Teeter often has specials on **wine** and **cheese**. Their focus is on promoting specific products while connecting to the customer through conversation about how to save time, money, and effort around subjects such as recipes or dinners.

While some of the results were expected, cluster 4 yielded surprising results. Originally, cluster 4 was interpreted to be all about food. However, after reviewing several terms through the filter viewer, it became clear that the cluster is about Harris Teeter promotions. Harris Teeter promotes various offers and customers respond to them because of the interests they have as can be seen in the other three clusters. Customers want to save money, receive great coupons, and maybe even enter sweepstakes and win free items. They also expect great customer service and a pleasant in-store experience. Since Express Lane is

increasing in popularity, this is a specific topic of interest. The four different clusters give us a good picture of Harris Teeter's Facebook presence and the interactions customers have with the company. Customers like to praise the aspects of Express Lane they like and complain about those they don't.

Negative Facebook Comments

Because 80% of the data had either positive or neutral sentiment, clustering for positive and neutral comments produced similar results as clustering for all of the data. While there were only 213 negative comments in total, the team felt that analyzing these would reveal negative perceptions people have about Harris Teeter.

The best model contained four clusters, high SVD resolution, and ran through the hierarchical clustering algorithm. While the clusters were not evenly distributed, the negative comments showed some patterns about what customers post about. Each cluster revolved around customers who had complaints about certain products, departments, or issues. A large portion of Harris Teeter sponsored Facebook posts were categorized as negative sentiment and were scattered throughout each cluster.

Cluster	Descriptive Terms	Conclusions
1. Poor Customer Service (54%)	Check, location, miss, items, price, service, customer, card, first, NC	People had issues with employees not being helpful with checking prices or item location.
2. Cakes & alternative locations (25%)	Lower, organic, prices, Harris, Teeter, bought, shop, cake, week, day	This cluster was about various issues customers had with cakes and various comments about going to shop at a different store next time.
3. Food Item Complaints (11%)	Bad, day, look, big, chicken, donate, fight, help, national, ingredients	This cluster is about complaints about chicken and various ingredients in foods. Comments posted by Harris Teeter about donations made it into this cluster as well.
4. Bad Experiences (9%)	Fix, pharmacy, spend, work, express, first, guys, horrible, lane, local	There are complaints about bad experiences at local stores in this cluster and issues related to the pharmacy and Express Lane.

Customer service is an issue when customers have a strong reaction to a bad experience or with specific products. While these results are not surprising, it is interesting that the first cluster is by far the largest and specifically points to North Carolina despite several states being represented in the data. This may suggest that North Carolina stores tend to have more complains about their customer service

department than other states. This issue can be explored further by first looking at the distribution of Harris Teeter stores throughout the southeast. If North Carolina is not especially saturated with Harris Teeter locations in comparison to other states, we could recommend investigating the cause of increased complaints about the customer service department.

Recommendations & Limitations

Based on our extensive analysis, we have developed two major categories of recommendations for Harris Teeter to consider. The first involves integrating the Express Lane application in-house and would address numerous issues identified during our research. The second category involves marketing Express Lane in order to increase participation and encourage new users.

Manage Express Lane Internally

Our first recommendation is for Harris Teeter to modify or end their existing relationship with the vendor managing the Express Lane customer interface. Internal insight provided by Ken Hayden and Chase Cabanillas allowed us to examine the data with more focus, providing context to interpret the results of our analysis. Because Harris Teeter does not manage the program internally, they have limited access to customer data, specifically demographic profiles. This hinders them from being able to narrowly segment customers through the Express Lane app in order to provide highly tailored offerings. The mobile functionality of the service gives Harris Teeter the opportunity to increase their customer reach and appeal to a wider target market, regardless of a customer's preferred shopping method.

Integrating the Express Lane app to in-house operations would also address several issues the team identified during analysis. Data from the survey, Facebook text clustering, and Express Lane specific analysis revealed coupons and promotions to be one of the most important attributes of grocery shopping. Harris Teeter customers in particular place high value on sales and VIC card offers. We suggest that Harris Teeter improve the Express Lane functionality so that coupon discounts can be applied at the time of purchase. Currently, customers are unable to use coupons on purchases made online, requiring them to take a raincheck and redeem the discount on their next eligible purchase. This is inconvenient for shoppers and may discourage them from using the service or even encourage them to select a different

retailer in the future. Our survey revealed that numerous shoppers perceived online shopping to be too expensive, so improving coupon redemption within Express Lane could help counteract this perception.

Similarly, promotions offered solely to online shoppers is another area we feel that Harris Teeter could improve. According to a study on global retailing by PricewaterhouseCoopers, researchers found that 91% of respondents were members of a retailer loyalty program, with 70% indicating that member only discounts or promotions are a top benefit (PWC, 2016). Our survey analysis reflected the study, indicating that of shoppers who preferred to shop in-store, 20% of them did so because they like to take advantage of in-store deals. Furthermore, Express Lane specific text clustering and topic analysis suggests that promotions are frequently mentioned in the data we reviewed, particularly those offered around holidays and for parties. It is likely that shoppers will utilize the online shopping feature when planning for large events, so money-saving incentives would help promote use of the service.

Next, by gaining control of the app, Harris Teeter could adjust the payment functionality to better serve customers. The team identified the need to expand online payment options, as they are currently limited. Although the web-based version of Express Lane allows for payment prior to pickup, orders made through the app require customers to go in-store and wait in line to complete the transaction. Hayden referenced complaints the company has received about the extra step required for orders made through the app. Aside from his insight and our analysis, we received testimony from a working professional that is a soon-to-be-father. He confirmed our findings and emphasized how beneficial the addition would be. Adjusting payment options would not only make Harris Teeter more shopper oriented, it would also lead to the creation and storage of transactional customer data, providing Harris Teeter with information beyond what products a customer buys.

The Express Lane interface both online and through the app lack key product categories - *organic* and *sale* items. Organic food is in demand across all retailing channels. As a result, we suggest that Harris Teeter adds a separate option for *organic* within the starting menu of the online shopping website. Doing so would prevent consumers from having to search for organic items. Additionally, because *sales* is a prominent term and topic from the survey, clustering, and sentiment analysis, we recommend adding a

separate category for sale items as shown below.

Lastly, it would be beneficial for Harris Teeter to integrate the Express Lane app and web-based platform to the inventory system for real-time product availability. Doing so would provide two primary benefits, the first of which is reducing the instances of stock out. Hayden stated that there have been situations in which a customer has selected an item through Express Lane, placed the order, and arrived at the store only to be informed that an item is not in stock. Similarly, the sentiment analysis clusters for Facebook posts about Express Lane concluded that customers were upset about promotions not being offered uniformly across all stores. Synchronizing inventory information would decrease variability, while promoting consistency across all of the store's retailing channels. It would also allow for localized pricing within specific Harris Teeter stores should price differentials exist among the various locations.

While we realize the financial implication and complexity of developing the infrastructure necessary to support the Express Lane platform internally, access to data is imperative, particularly for industries marked by intense competition such as the grocery retailing business. The driving force behind this multi-step recommendation is to exceed expectations and delight the company's most profitable

customers. The long-term benefits are significant and therefore should be addressed within the relationship Harris Teeter currently has with the third party vendor managing the Express Lane customer interface. To remain competitive, it's important the Harris Teeter focus on transactional behavior of their customers to better serve them.

Express Lane Marketing Efforts

Our second recommendation is for Harris Teeter to increase marketing efforts around Express Lane. We feel that Harris Teeter needs to establish non-price based reasons for customers to select them as their preferred grocery retailer and to remain competitive. Social media in particular is an inexpensive and easy way to increase marketing to online audiences. The retailer has a relatively prominent presence online, and is moderately active across it's various social media sites. However, it is not enough to simply increase the frequency of posts. In general, Harris Teeter should post advertisements about directly relating to Express Lane. Doing so would increase awareness of the service and act as a reminder for current subscribers. Specifically, Harris Teeter should increase posts with content or paid promotions for organic products to highlight the company as a go-to for organic food. Adoption of both actions will more effectively engage with customers through their preferred channel, which will improve customer loyalty and drive growth to increase the bottom line.

Additionally, we suggest that the company emphasizes the quality of produce to Express Lane shoppers through their marketing efforts. Survey results indicated that *availability of fresh produce* is the most important attribute for online grocery shopping. Not surprisingly, *fresh food* is also one of the main concerns of Harris Teeter customers that we identified, both in-store and online. Analysis showed that in-store customers are mostly content with the freshness of food, represented by the positive sentiment of Harris Teeter reviews on Yelp and Facebook. However, this perception is not shared by online customers. Harris Teeter should take steps to ensure the supply of fresh produce to online customers. One method could be granting online orders priority in selecting produce from incoming shipments or selection from a separate batch entirely, so that Express Lane users have the freshest options. Positive perception is key for

achieving customer satisfaction, even more so than in-store shoppers. It is also important to note that this is not a marketing specific task, as logistics would obviously need to be involved to implementation.

Limitations

There are four main limitations that must be addressed in order to provide a fair overview of the results and their applicability to the business problems the team sought to address.

First, there was limited availability in Harris Teeter-specific demographic data. The survey sample participants were not necessarily Harris Teeter customers and there was not a survey question addressing which grocery store, if any, respondents primarily shopped. Although Harris Teeter provided some basic information for the project, specific demographic information relating to their online shopping demographic profile could not be provided due to the Express Lane customer interface being managed by a third party.

Second, the survey sample is subject to voluntary-response bias, as well as those associated with convenience samples. Because the team distributed the survey online through social media and email, respondents who chose to participate were primarily personal contacts, and therefore from demographic groups similar to the team. The results contain unequal, non-random representation of demographic profiles, so insights developed assume that people who did not participate would respond to questions in a similar way. Furthermore, there may not be equal representation of satisfied and dissatisfied customers. The voluntary sample may be a poor match of the population we set out to understand. Inferences made from the survey data are not as reliable as a truly random sample would have produced, nor are they statistically relevant. Additionally, data selected for text mining originated from multiple platforms, each with a unique set of users. For example, Yelp data consisted solely of reviews for Harris Teeter stores in Charlotte, whereas Facebook data is national. The data should not necessarily be combined or compared, however, it seems appropriate given the scope of our research study.

The final limitation we must consider is the *Express Lane* mentions in the Yelp and Facebook data sets, which may be used interchangeably with the term *express checkout*. In the relatively small data samples the team used, we were able to check enough of our results to be comfortable to move forward

with analysis. However, it is possible that we may have missed something. In a larger data set, this would

be an important limitation to address through more detailed text categorization.

References

Annual Report. (2013). Harris Teeter Supermarkets, Inc. Retrieved from https://www.last10k.com/sec-filings/85704

Cabanillas, Chase. Personal Communication. April 2016.

Charlotte Observer Staff. (2016). Grocery Wars: Harris Teeter takes top spot, but Walmart still going strong. Retrieved from http://www.wbtv.com/story/31682932/grocery-wars-harris-teeter-takes-top-spot-but-wal-mart-still-going-strong

Goldstein, Steve. (2015). Here's why restaurants are eating Wal-Mart's, Whole Foods lunch. MarketWatch. http://www.marketwatch.com/story/from-wal-mart-to-whole-foods-people-are-choosing-eating-out-over-eating-in-2015-10-15

Halzack, Sarah. (2015). The Staggering Challenges of the Online Grocery Business. The Washington Post. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2015/01/20/the-staggering-challenges-of-the-online-grocery-business/

Hayden, Ken. Personal Communication. March 1st, 2016.

Martin, Jenna. (April 27, 2016). Hometown grocer notches win in market-share battle. Retrieved from http://www.bizjournals.com/charlotte/news/2016/04/27/hometown-grocer-notches-win-in-market-share-battle.html.

Peralta, Katherine. (2016). Grocery Wars: Harris Teeter takes top spot, but Walmart still going strong. *Charlotte Observer*. Retrieved from http://www.charlotteobserver.com/news/business/article70779657.html

Peralta, Katherine. (2015). Buying groceries online: A niche market or the future of shopping. Charlotte Observer. Retrieved from http://www.charlotteobserver.com/news/business/article39358632.html

Peralta, Katherine. (2015). Publix grocery delivery service Shipt to launch in Charlotte Sept. 29. Charlotte Observer. Retrieved from http://www.charlotteobserver.com/news/business/biz-columns-blogs/whats-in-store/article36085020.html

Peralta, Katherine. (2015). Walmart, Harris Teeter still on top of local grocery scene, but competition growing. Charlotte Observer. Retrieved from http://www.charlotteobserver.com/news/business/article17232803.html

Peralta, Katherine. (Sept 10, 2015). Lidl heats up Charlotte's low-cost grocery competition. Charlotte Observer. Retrieved from http://www.charlotteobserver.com/news/business/article34643982.html

Pricewaterhouse Coopers. (February 2016). Total Retail 2016. Retrieved from http://www.pwc.com/us/en/retail-consumer/publications/assets/Total-Retail-Global-Report.pdf

Smith, Diana. (2015). Grocery Retailing - US - November 2015. Mintel Academic. Retrieved from http://academic.mintel.com.librarylink.uncc.edu/display/716858/

Smith, Diana. (2015). Grocery Retailing Infographic Overview. Mintel Group.

Springer, Jon. (2016). Study: online grocery shopping no longer a niche. *Supermarket News*. Retrieved from http://supermarketnews.com/new-consumer/study-online-grocery-shopping-no-longer-niche

Womack, Brian. (2015). Google to Start Testing Grocery Delivery This Year. Bloomberg Business. Retrieved from http://www.bloomberg.com/news/articles/2015-09-08/google-to-start-testing-grocery-deliveries-this-year

Hoovers. (N.D.) Grocery Stores and Supermarkets Report Summary. Retrieved from http://www.hoovers.com/industry-facts.grocery-stores-supermarkets.1535.html

APPENDIX A
Survey Analysis

Figure A1 - Harris Teeter Locations in Charlotte
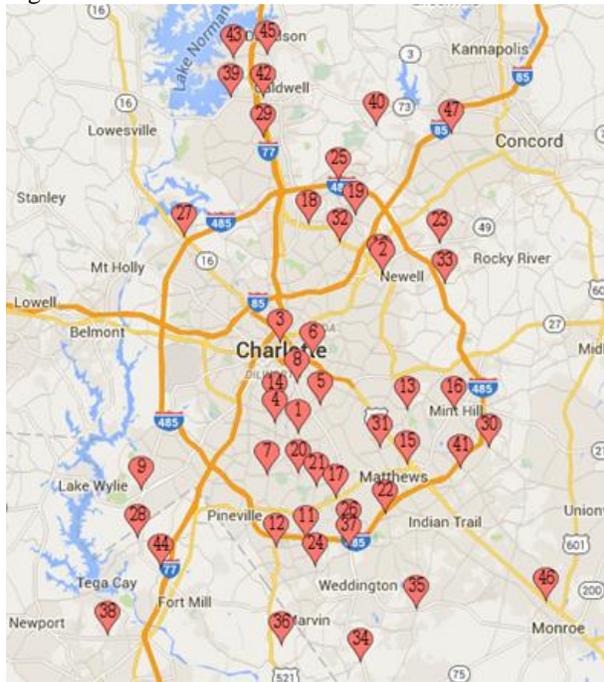
Table A1 - Survey Regression

Parameter Estimates					
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2.0535	1.0053	Infty	2.04	0.0411
RetailChoice_BrandLo	1.4180	1.0307	Infty	1.38	0.1689
RetailChoice_LowerPr	-1.2261	1.0597	Infty	-1.16	0.2473
RetailChoice_FreeShi	-1.6313	1.0786	Infty	-1.51	0.1304
RetailChoice_PickupD	-0.4797	1.2326	Infty	-0.39	0.6971
RetailChoice_Reviews	1.0984	0.8328	Infty	1.32	0.1872
RetailChoice_EaseofL	-1.6361	1.4765	Infty	-1.11	0.2678
RetailChoice_Product	0.9622	1.0263	Infty	0.94	0.3485
RetailChoice_Rewards	13.9299	722.87	Infty	0.02	0.9846
RetailChoice_Limited	0.2374	1.1663	Infty	0.20	0.8387
RetailChoice_Multich	14.1975	820.72	Infty	0.02	0.9862
RetailChoice_RangePr	0.2910	1.1190	Infty	0.26	0.7948
RetailChoice_EasyWeb	0.3925	0.9621	Infty	0.41	0.6833
RetailChoice_MobileF	0.3046	1.1180	Infty	0.27	0.7853
RetailChoice_Coupons	0.9854	0.9244	Infty	1.07	0.2865
RetailChoice_EasyRet	0.2251	0.8950	Infty	0.25	0.8014
RetailChoice_Reputat	0.5429	1.1644	Infty	0.47	0.6410

Table A2 - Survey Regression

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
		Parameter Estimates			
Intercept	-3.4767	1318.11	Infty	-0.00	0.9979
Gender	0.4788	0.8862	Infty	0.54	0.5890
Age	-0.05885	0.3117	Infty	-0.19	0.8502
Household_	1.0220	0.6831	Infty	1.50	0.1346
ChildrenUnder18	2.4161	94.3667	Infty	0.03	0.9796
ChildrenUnder5	11.3155	708.12	Infty	0.02	0.9873
Children5to12	6.3174	841.25	Infty	0.01	0.9940
Children13plus	-2.4526	698.48	Infty	-0.00	0.9972
ChildrenNone	-13.0763	654.60	Infty	-0.02	0.9841
IncomeLevel1	13.4262	1044.40	Infty	0.01	0.9897
IncomeLevel2	2.2632	1088.23	Infty	0.00	0.9983
IncomeLevel3	1.3303	1088.23	Infty	0.00	0.9990
IncomeLevel4	1.2081	1088.23	Infty	0.00	0.9991
IncomeLevel5	1.1699	1088.23	Infty	0.00	0.9991
IncomeLevel6	12.5086	1186.07	Infty	0.01	0.9916
IncomeLevel7	-0.9497	1088.23	Infty	-0.00	0.9993
IncomeLevel8	13.7452	1508.32	Infty	0.01	0.9927
IncomeLevel9	-0.2900	1088.23	Infty	-0.00	0.9998
IncomeLevel10	12.1092	1172.87	Infty	0.01	0.9918

Table A3 - Survey Regression

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
		Parameter Estimates			
Intercept	3.3002	1.1335	Infty	2.91	0.0036
Shop_Retail	0.04987	0.8873	Infty	0.06	0.9552
Shop_Books	-0.2245	0.7892	Infty	-0.28	0.7760
Shop_Jewelry	-0.1292	0.8937	Infty	-0.14	0.8851
Shop_HealthBeauty	0.1559	0.7951	Infty	0.20	0.8445
Shop_Toys	-1.4010	0.9090	Infty	-1.54	0.1233
Shop_SportingGoods	-0.2449	0.8977	Infty	-0.27	0.7850
Shop_HomeFurniture	-0.8968	0.8418	Infty	-1.07	0.2867
Shop_Electronics	-0.09302	1.0162	Infty	-0.09	0.9271
Shop_Appliances	-0.8391	0.8603	Infty	-0.98	0.3294
Shop_Other	0.02235	1.1027	Infty	0.02	0.9838

Table A4 - Survey Regression

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
		Parameter Estimates			
Intercept	6.2320	2.2211	Infty	2.81	0.0050
WhyGrocOnline_PriceC	-1.4473	1.3983	Infty	-1.04	0.3006
WhyGrocOnline_Conven	-1.5781	1.6964	Infty	-0.93	0.3522
WhyGrocOnline_Specif	18.1067	402.16	Infty	0.05	0.9641
WhyGrocOnline_Smalle	-3.5899	2.2434	Infty	-1.60	0.1096
WhyGrocOnline_MajorT	11.9070	1044.40	Infty	0.01	0.9909
WhyGrocOnline_Subscr	-2.0702	1.5155	Infty	-1.37	0.1719
WhyGrocOnline_Coupon	-3.7143	2.2613	Infty	-1.64	0.1005
WhyGrocOnline_Budget	4.5646	2.5599	Infty	1.78	0.0746
WhyGrocOnline_SaveTi	-2.8925	1.8681	Infty	-1.55	0.1215
WhyGrocOnline_AvoidC	-2.7071	1.3105	Infty	-2.07	0.0389
WhyGrocOnline_247Acc	3.5985	2.3580	Infty	1.53	0.1270

APPENDIX B
Yelp Analysis
Table B1 - Yelp Data Cluster

Cluster	Descriptive Terms	Frequency	Percentage	RMSSTD
1	+lack +'park lot' +park +locate +space +park +bag +stop charlotte last decent right quality +long home +people +buy +visit ht	89	16%	0.130202023
2	+bar +customer +fresh +meat +service always +friendly helpful +love +salad +day teeter +employee produce staff great harris +want back +great	359	64%	0.127791918
3	+high +item +sale +price +grocery +feel +cheese +shop +'grocery store' +location +small +big ht first +area +find food +good +store charlotte	114	20%	0.137523417

Table B2 - Yelp Data Cluster

Cluster ID	Descriptive Terms		Frequency	Percentage
1	+order +lunch +walk +cheese back +want +check +counter +offer 'customer service' +service +buy +bar +customer ht +bag hot +day +time +look	...	131	23%
2	+lack +bag +'salad bar' hot +forget +salad quality south +bar +variety expensive charlotte +market +meal +pay decent +long food +checkout +food	...	43	8%
3	+nice +price +item teeters +selection +shop produce +clean +large +high +friendly +grocery staff +small +good +find +section +wine pretty +store	...	194	35%
4	+'park lot' +entrance +lot +park +space +area starbucks +aisle +location +big right +line +small first +little +stock well pretty +walk +few	...	85	15%
5	+customer +service helpful 'customer service' always +employee +experience harris teeter +love +cashier +great great +work +friendly staff +stock +offer well +clean	...	109	19%

Table B3 - Yelp Data Cluster

Cluster ID	Descriptive Terms		Frequency	Percentage	
1	+'salad bar' +bar +salad hot +pizza sushi +lunch +fresh starbucks +area +work +great +cheese +food +want	...	69	12%	
2	'customer service' +customer +service +employee +cashier +manager +experience back +day +work +know harris +item +want +line	...	74	13%	
3	+nice +selection +clean +friendly staff produce +wine +price always teeter pretty +item harris +grocery +shop	...	251	45%	
4	+lack +wish +bag +locate quality charlotte decent +checkout stop home +high +visit food right +buy	...	53	9%	
5	online +'online shop' money +card vic +shop +employee +week great +service +time +love +find +bag home	...	31	6%	
6	+cheese +offer +want back +manager +lunch +lot right +people +card +location money +meat +day +park	...	84	15%	

Table B4 - Yelp Data Cluster

Cluster ID	Descriptive Terms		Frequency	Percentage
1	+card down +old +run +open +move +look +experience +few +beer +cashier +buy great +aisle staff	...	19	3%
2	sushi today +lunch +chicken +pizza hot +bar +cheese +buy +walk +offer +want +manager ht +place	...	60	11%
3	+'salad bar' +bar +salad hot +lunch +fresh +meat starbucks +great +pizza +love +wine +beer +food +section	...	55	10%
4	+friendly helpful staff +price +clean produce always great +great +selection teeter +shop +nice harris +find	...	136	24%
5	'customer service' +customer +service +cashier +manager +employee back +experience +day +offer harris +know +order +little +item	...	72	13%
6	+park +lot +'park lot' +small ht pretty +big +aisle +nice +store +location +selection +old +area down	...	145	26%
7	+'check out' +'decent hot meal' +'difficult time find' +'expensive state' +barrier +bump +follow +lack +bag +driver +'great place' +charge +competitor +feeling quality	...	12	2%
8	+order +walk +option +meat +offer +day +cheese +employee +want +card +food +time back +shop +manager	...	63	11%

Table B5 - Yelp.com Logistic Regression
  1. Negative dummy

```
      Likelihood Ratio Test for Global Null Hypothesis: BETA=0

       -2 Log Likelihood            Likelihood
     Intercept     Intercept &        Ratio
        Only       Covariates       Chi-Square       DF      Pr > ChiSq

       363.255        309.861        53.3938          6        <.0001


                    Analysis of Maximum Likelihood Estimates

                                    Standard       Wald                 Standardized
     Parameter       DF    Estimate    Error   Chi-Square   Pr > ChiSq     Estimate    Exp(Est)

     Intercept        1     2.0216    0.8532       5.61       0.0178                     7.551
     TextTopic_raw1   1     2.2002    0.8608       6.53       0.0106       0.1828        9.027
     TextTopic_raw2   1    -1.9075    1.1020       3.00       0.0835      -0.1805        0.148
     TextTopic_raw5   1    -3.9032    2.0525       3.62       0.0572      -0.2450        0.020
     TextTopic_raw6   1    -3.1667    1.7980       3.10       0.0782      -0.1906        0.042
     price            1    -0.7467    0.3615       4.27       0.0389      -0.2393        0.474
     rating           1    -0.6263    0.1175      28.41      <.0001       -0.4285        0.535
```

  2. Positive dummy

```
Intercept  TextTopic_raw2  TextTopic_raw3  TextTopic_raw5  price  rating


      Likelihood Ratio Test for Global Null Hypothesis: BETA=0

       -2 Log Likelihood            Likelihood
     Intercept     Intercept &        Ratio
        Only       Covariates       Chi-Square       DF      Pr > ChiSq

       578.629        480.079        98.5492          5        <.0001


                    Analysis of Maximum Likelihood Estimates

                                    Standard       Wald                 Standardized
     Parameter       DF    Estimate    Error   Chi-Square   Pr > ChiSq     Estimate    Exp(Est)

     Intercept        1     2.3905    0.6705      12.71       0.0004                    10.918
     TextTopic_raw2   1     1.2546    0.6684       3.52       0.0605       0.1187        3.506
     TextTopic_raw3   1     1.7814    0.8747       4.15       0.0417       0.1362        5.938
     TextTopic_raw5   1     1.9395    1.0720       3.27       0.0704       0.1217        6.955
     price            1    -1.7335    0.2944      34.66      <.0001       -0.5556        0.177
     rating           1     0.4847    0.0953      25.84      <.0001        0.3316        1.624
```

APPENDIX C
Express Lane Specific Analysis - Yelp & Facebook
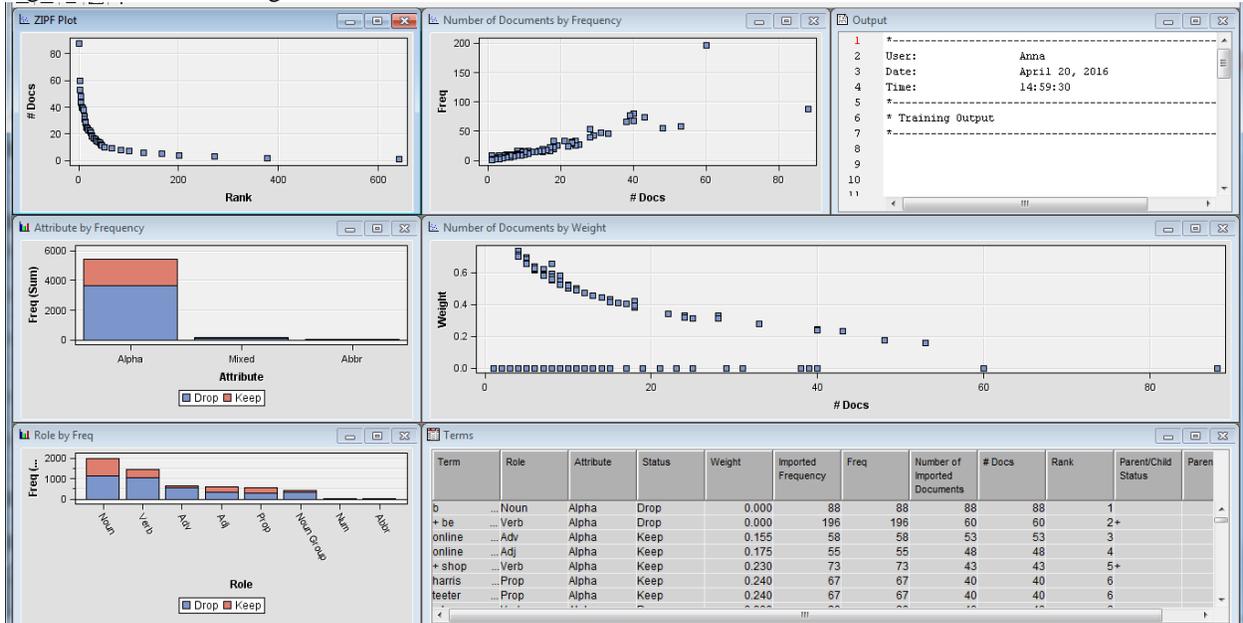
Figure D1 - Text Filtering Results

Figure D2 - Pissed Consumer Sentiment Analysis

41 Negative, 14 Neutral, 52 Positive