ABSTRACT

LU, JINZHAO. Engaging Omni-Channel Consumers During Purchase Decisions. (Under the direction of Dr. Trevor J. Little).

Previously, bricks and mortar stores enabled consumer to touch and feel the products. Today, Internet e-commerce retailers attract consumers with more product choices, lower prices, and more detailed product information. Increasing numbers of consumers empowered with digital technologies, constantly search information and compare prices online, experience products in bricks and mortar stores, and make final purchase through different channels. The distinct boundaries across different shopping channels are vanishing. Specifically, consumers switch between different channels and touch points constantly, interchangeably, and simultaneously, who are considered as omni-channel consumers. Firms are trying to fully capture the enormity of the digital opportunity and better understand the consumer’s journey to purchase, especially, at what point along the journey, the customers can be easily influenced, and how to interact, especially given the consumer the supporting technologies such as smart phones and tablets that are readily available to the consumer.

This study developed and validated a touch points based customer experience (TPE) model in today’s omni-channel retailing environment. This research studied the impact of consumers channel choice (single channel VS multichannel) on the proposed TPE model and narrowed down the scope to a specific touch point – communicative touch point and analyzed the raw corporation data from a global 500 specialty retailer. In specific, this research investigated the role of RFM (Recency, Frequency, Monetary) on predicting omni-channel consumers’ purchase probability for targeted promotions.
Engaging Omni-Channel Consumers During Purchase Decisions

by
Jinzhao Lu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Textile Technology Management

Raleigh, North Carolina

2017

APPROVED BY:

___________________________  __________________________
Dr. Trevor J. Little        Dr. Yingjiao Xu
Chair of Advisory Committee

___________________________  __________________________
Dr. Marguerite Moore       Dr. Jonathan Bohlman
DEDICATION

I dedicate this dissertation and give special thanks to my wonderful family. Thank you for your invaluable support throughout my entire doctorate program. This dissertation is also dedicated to my mentor, Dr. Trevor J. Little for his support and guidance.
BIOGRAPHY

Jinzhao Lu joined the 3+X program, which is a joint accelerated Bachelor’s and Master’s program between Donghua University in China and North Carolina State University, Raleigh in 2012. He received his bachelor’s degree of Textile Engineering from Donghua University in July 2013. He received his Master of Science degree majoring in Textiles and minoring Statistics in 2014. Then he continued pursuing his Doctorate degree in Textile Technology and Management in the College of Textiles, North Carolina State University, Raleigh, NC. His research interests focus on omni-channel consumer insights, data mining and data analytics, and supply chain analytics.
ACKNOWLEDGMENTS

First, I would like to express sincere gratitude and thanks to my mentor, Dr. Trevor J. Little, for his consistent, patient and generous guidance and supervision of my research from the beginning to the end. His advice on my research as well as on my future career was inspiring and encouraging.

Second, I am grateful to my committee members, Dr. Jonathan Bohlmann, Dr. Yingjiao Xu and Dr. Marguerite Moore, for their insightful comments during the course of my research. They are great professors who incited me to widen my research from various perspectives and encouraged me to foster my critical thinking ability.

Third, I would like to thank Dr. Cole Conley, Dr. Matt Moyer, Dr. Liangyu Zhu, and Dr Zhentao Tong from Department of Statistics, North Carolina State University. They provided me with great consulting help during data cleaning and statistical modeling process.

My final appreciation goes to my beloved family. My parents always teach me to work with determination, become a useful person, and be always respectful to others. Their support and love is the greatest motivation of my study and life here in United States.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF EQUATIONS</td>
<td>ix</td>
</tr>
<tr>
<td><strong>CHAPTER 1</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Research objectives</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Dissertation structure</td>
<td>4</td>
</tr>
<tr>
<td><strong>CHAPTER 2</strong></td>
<td></td>
</tr>
<tr>
<td>Literature Review</td>
<td>5</td>
</tr>
<tr>
<td>Part A. Omni-channel Consumer Behavior Insights</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Consumer behavior definition</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Online consumer behavior</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Multi-channel consumer behavior</td>
<td>8</td>
</tr>
<tr>
<td>2.4 Consumer decision-making process and touch points</td>
<td>12</td>
</tr>
<tr>
<td>2.5 Customer experience</td>
<td>17</td>
</tr>
<tr>
<td>2.6 Customer experience measurements</td>
<td>18</td>
</tr>
<tr>
<td>2.7 Omni-channel consumer behavior</td>
<td>27</td>
</tr>
<tr>
<td>2.8 Theoretical Framework</td>
<td>31</td>
</tr>
<tr>
<td>Part B. Omni-Channel Consumers Engagement Strategies</td>
<td>32</td>
</tr>
<tr>
<td>2.9 Consumer data model</td>
<td>33</td>
</tr>
<tr>
<td>a. Consumer segmentation</td>
<td>34</td>
</tr>
<tr>
<td>b. Recommendation engine</td>
<td>35</td>
</tr>
<tr>
<td>c. Cross selling, up selling</td>
<td>37</td>
</tr>
<tr>
<td>d. Predictive analytics</td>
<td>39</td>
</tr>
<tr>
<td>e. Sentiment analysis</td>
<td>41</td>
</tr>
<tr>
<td>f. Location based marketing</td>
<td>42</td>
</tr>
<tr>
<td>g. In-store digitalization</td>
<td>44</td>
</tr>
<tr>
<td>h. Tools and algorithms</td>
<td>46</td>
</tr>
<tr>
<td>i. CLV (customer lifetime value) and RFM (recency, frequency, monetary)</td>
<td>47</td>
</tr>
<tr>
<td><strong>CHAPTER 3</strong></td>
<td></td>
</tr>
<tr>
<td>Research Methodology</td>
<td>54</td>
</tr>
<tr>
<td>3.1 Data collection</td>
<td>54</td>
</tr>
<tr>
<td>3.1.1 Research Objectives 1-8 data collection: Survey questionnaire</td>
<td>54</td>
</tr>
<tr>
<td>3.1.2 Research Objective 9 data collection: Wharton School dataset</td>
<td>54</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. Omni-channel consumer journey model .................................................. 12
Figure 2. Consumer decision making model .............................................................. 13
Figure 3. Dynamic Consumer Decision Journey ...................................................... 16
Figure 4. Touch point based experience (TPE) theoretical framework for objective 1-8.... 31
Figure 5. Consumer interaction data ........................................................................ 34
Figure 6. Omni-channel consumer engagement review graphic model ..................... 47
Figure 7: Comparison of overall generic emails and targeted emails effects ............. 98
Figure 8: Distribution of estimated fixed effect for Recency ...................................... 100
Figure 9: Distribution of estimated fixed effect for Frequency ................................... 101
Figure 10: Distribution of estimated fixed effect for Monetary ................................. 103
Figure 11: Variance component from the random effect of monetary ....................... 104
Figure 12: Histogram of variance component from the random effect of monetary ...... 105
Figure 13: ROC Curve for GLMM and Baseline Logistic Regression ....................... 107
Figure 14: Touch points based omni-channel experience results for all customers ....... 111
Figure 15: Touch points based omni-channel experience: Single vs Multi-channel ...... 113
LIST OF TABLES

Table 1. Descriptive statistic of gender and shopping channel........................................75
Table 2. Summary of Exploratory Factor Analysis Results .............................................78
Table 3. CFA model fit indices .......................................................................................81
Table 4. CFA validity results ..........................................................................................83
Table 5. Factor correlation matrix with square root of the AVE on bold number ..........84
Table 6. SEM model fit ................................................................................................85
Table 7. SEM path analysis results ................................................................................87
Table 8. Test of invariance results .................................................................................90
Table 9. Multi-group moderation analysis results .........................................................92
Table 10. Parameter estimates with all emails ..............................................................94
Table 11. Purchase probability given the receipt values of all emails .........................94
Table 12. Parameter Estimates with targeted emails only ............................................96
Table 13. Purchase probability given the receipt values of targeted emails .................97
LIST OF EQUATIONS

Equation 1. Customer lifetime value calculation 1 ...................................................... 48
Equation 2. Customer lifetime value calculation 2 ...................................................... 49
Equation 3. CLV calculation based on RFM ................................................................. 51
Equation 4. Logistic Function modeling targeted emails’ effects .................................. 63
Equation 5. RFM normalization function .................................................................... 69
Equation 6. Generalized linear mixed effect model ..................................................... 70
Equation 7. Generalized linear mixed effect model with random effect ....................... 71
Equation 8. Simple logistic model with fixed effect ..................................................... 73
CHAPTER 1

Introduction

1.1 Background

“The seeker embarks on a journey to find what he wants and discovers along the way, what he needs.” - Wally Lamb (2008) from The Hour I First Believed

Widespread adoption of different shopping channels by all US population has brought about new opportunities for businesses to accurately inform the right consumers about product availability, features and special offers. A survey from Economistinsights and Mckinsey (Edelman & Banfi, 2014) shows that 44% of shoppers use their mobile phone to search promotions, compare prices, and check reviews, and 72% of companies believe their budget for customer insights is too low while 45% of companies admit they have limitation to understand how their consumer interact with them digitally. Well-known retailers such as Walmart, Target, and Macy’s are reducing workforce to become more agile, effective organizations that can react to changes in customer behavior; especially the shift to digital shopping. Target will invest particularly on areas such as digital, personalization, data and analytics, and engineering, as stated by its spokeswoman Molly Snyder (Wahba, 2015).

More and more, firms are trying to fully capture the enormity of the digital opportunity and better understand the consumer’s journey to purchase; especially the point along the journey when the customers are easily influenced about product choice. The intervention strategy
must be informative and technologically apt, especially given that the consumer’s supporting technologies are smart phones, tablets, and increasingly a wide range of apps.

Retail, especially omni-channel, has complicated the consumer decision process and coupled with advancing technology availability, retailers are working diligently to respond to the changing marketplace. In particular, planning and forecasting are experiencing challenges in assortment planning at the SKU level of detail, which can lead to obsolete merchandise. The shelf-life of many fashion items is sufficiently short that new design and development inputs must be accurately predicted to be able to operate a business successfully.

Big data analysis, such as predictive analytics has been advocated as one of the possible approaches to be better able to forecast consumer behavior. Retail has access to ‘big data’ but lacks many of the smart algorithms that can take into consideration the many factors that influence a consumer decision at the point of purchase. The risk of making bad business decisions based on novel algorithms or poor data analysis is exponentially increasing. In addition, effective means of communication with the consumer continually changes as more and more psychographic and demographic factors are investigated.

This research will adopt an Omni-channel Consumers Engagement perspective that includes, among other things, data and analytics, promotional and marketing approaches, omni-channel strategies and the relationships between touch points and consumer behavior to find more accurate ways – when, where, and how to interact with omni-channel customers.
1.2 Research objectives

A study conducted by IBM involving more than 1500 retail CEOs, 93 percent of whom considered it a top priority to better understand, predict and give customers what they want. The management of retailing and marketing has developed toward building a seamless consumer shopping experience, where the right products are shown at the right place, also the timely and relevant communications are sent to the right channel (IBM, 2011). There are three main objectives of the study.

1. The first objective is to determine consumer experience dimensions from a new perspective: the consumer’s shopping journey touch points, which is the moment when customers interact with the product, service, or brand. This study developed and validated a touch point based customer experience (TPE) model in today’s omni-channel retailing environment. Particularly, seven touch points include atmospheric, communicative, product, technological, process, employee-customer, customer-customer were examined.

2. The second objective is to examine the relationship between consumers’ single touch point during decision making process and consumers’ overall shopping experiences. Also, this study investigates the impact of consumers channel choice (single channel VS multichannel) on the proposed TPE model.

3. The third objective is to leverage real world customer touch point data, and build, apply and evaluate the RFM (Recency, Frequency, Monetary) models in predicting omni-channel consumers’ purchase probability during targeted promotions.
1.3 Dissertation structure

There are four major parts of the dissertation. Chapter Two first presents a review of previous studies on the consumer behavior, and how consumer behavior may have changed in today’s omni-channel retailing environment. Customer experience and its dimensions are also discussed. Objectives of the research were stated along with the literature review. Nine research objectives are proposed based on previous literature. Then, current consumer analytics marketing strategies and the major algorithms behind these market strategies are summarized as a review in this subject. Chapter Three presents data collection, data cleaning, data instrument and data analytics methodologies. Research Objectives 1-8 are based on primary survey data, and Research Objective 9 is built on the secondary industry data. Thus the structure of the dissertation is split into two segmentations below. Chapter Four illustrates the results of Research Objectives 1-8 and Research Objective 9. Chapter Five discusses the results of the Research Objectives and makes conclusions. Lastly, implications, limitations and future study are discussed in Chapter Six.
CHAPTER 2

Literature Review

Part A. Omni-channel Consumer Behavior Insights

2.1 Consumer behavior definition

Consumer behavior is defined by Kotler as the mental, emotional and physical activity that consumers use during selection, purchase, use and disposal of products and services that satisfy their needs and desires (Kotler, 1999).

Schiffman and Kanuk (2010) describe the consumer behavior as the behavior that consumer show in searching, purchasing, using, evaluation and disposing of products and services that satisfy their needs or expectations. These behavior include what the customer buy, how they buy, why, when, where they buy, how they use the products, how they evaluate and dispose of it after purchasing, and how the experience influence the future behavior.

We define consumer behavior as a set of psychological and physical activities that the individual or organizational consumers display during need recognition, searching, comparing, purchasing, and post-purchasing of products or services.

The main application of studies in consumer behavior is to optimize the 5p’s of a marketing strategy. It was about mid 1950s when companies realized that in order to compete with other companies, they must understand the needs, wants or interests of a specific target markets and design, produce, merchandise the products or services that meets the consumer needs.
Through extensive market research or data analytics, companies now can better understand consumer behavior and launch marketing campaigns such as segmentation, targeting, positioning as well as marketing mix based on the consumer behavior. Other applications include using social marketing to get inspired by specific consumer groups rather than sell products, or making public policies based on the consumer behavior (USC, 2016).

2.2 Online consumer behavior

E-commerce was reported less than 1 percent in terms of total retail revenue in 1998, while several years later, the number has grown to about 11 percent in 2015 (Oh, 2015). With more digital technology developed and more personalized marketing information presented, it is convenient and efficient for consumers to search for relevant information and make final purchases online. There is no doubt that more and more dollars will stream to digital purchase in all sectors of the retailing industry.

During the past decade, researchers and marketers have studied several aspects of online consumer behavior. M. Coufario (2002) claims that online consumers have more power and are more demanding compared to offline consumers because online consumers could get more information online, from their friends, or from other products users (Gatautis & Kazakevičiūtė, 2012).

Many factors influence online consumer behavior such as demographic and geographic information, perceived risk, social influence, website designing, personal digital skills, as well as diffusion of technologies. Gong and Maddox (2011) proposed five factors that could
influence consumer online and offline shopping behavior, which are channel-risk perceptions, price-search intentions, search outcome, evaluation outcome, and delivery time.

A survey carried out by Price Waterhouse Cooper reveals that the top reason motivating consumer’s online shopping behavior is the convenience factor. The consumers are able to shop whenever they want and wherever they want. Other key factors that harbored by consumers are better price, more products choice, easier to compare price, products and offers, as well as free and faster delivery (PWC, 2015).

On the other hand, the biggest drawback of online shopping is that consumers are not be able touch, feel, or use the product before making a final purchase. This lack of face-to-face experience discourages some consumer to buy online and they still prefer to buy in brick-and-mortar stores (Delft, 2013).

However, even though researchers have unveiled the key motivations of online consumer behavior, retailers failed to fully utilize it in improving consumer satisfaction or optimizing marketing campaigns. For instance, according to a Price Waterhouse Cooper survey, by offering free shipping, retailers can increase margin by eight to twelve percent while nearly sixty percent of the retailers charge for shipping or handling fees (PWC, 2015).

A report from a Forrester Research survey also shows that shipping costs are the most common reason that cause consumer shopping cart abandonment. 75 percent of consumers may shift to another retailer if the shipping fee is too high or not free. Also, when free shipping becomes a type of promotion, it is the most effective way to drive sales during
holiday season, where consumers tend to spend 10 to 15 percent more in total order value. The retailers need to calculate whether this increase in revenue will cover the expense of a free shipping campaign (Fedex, 2011).

Another threshold that keeps the consumer shifting is the return fees and unclear return policy. Survey from comScore also indicates that the ease of making returns or exchanges is one of the most important factors during check out process (Wahba, 2015).

Until recently, more and more retailers launch free shipping offers. For example, to compete with Walmart, Target unveiled a new weapon in its holiday season e-commerce battle: free shipping and free returns, which contributes to its 20 percent increase in digital sales for the quarter (Garcia, 2016).

2.3 Multi-channel consumer behavior

Multichannel marketing refers to the marketing strategies using a combination of indirect and direct consumer interaction channels such as websites, retail stores, mail order catalogs, direct mail, email, mobile, etc. (SAS, 2016).

Multi-channel management was defined as the enhancement of customer experience through customer acquisition, retention and development by designing, deploying, and evaluating of different communication channels (Neslin et al., 2006).

Multi-channel consumers are usually engaged in three types of activities: 1) Shopping across at least two channels such as brick and mortar, online, or mobile applications. 2) Purchasing
products from same retailer but through more than one channel. 3) Leveraging different channels to make a single purchase (PWC, 2015). Researcher and marketers have previously dealt with consumer channels separately, and considered research shoppers as the consumer search in one channel and make purchase in another channel (Verhoef, Neslin, and Vroomen, 2007).

Retailers recently have finally understood that consumers do not realize the difference between different channels. What consumers care most is to find the desired products to meet their need with great value in term of money and price (Garcia, 2016).

Previously, bricks and mortar stores enabled consumer to touch and feel the product. On the other hand, internet e-commerce retailers tried to attract consumers with more product choices, lower prices, and more detailed product information. Nowadays, with more and more consumer empowered with digital technologies such as mobile and tablets, they constantly search information, compare prices in bricks and mortar stores. The distinct boundary across different shopping channels is vanishing. Specifically, the omni-channel consumers switch between different channels and touchpoints constantly, interchangeably, and simultaneously (Verhoef, Kannan & Inman, 2015).

Showrooming is a phenomenon that consumers visit local brick-and-mortar stores to check out products but then purchase the products online. A Survey shows (Khan, 2014) that over 50% of smartphone owners have researched price in a store, among which a third of these consumers end up purchase at a competitor. The most important reason why consumer
showroom is that other online vendors offer lower prices and consumers do not want to pay for any shipping or handling fees.

On the other hand, more and more consumers are webrooming, where they research products online and then go to the local stores to make a final purchase. Consumers prefer to touch and feel a product before purchasing and they don’t want wait for delivery of a product (Khan, 2014).

Under this consumer-driven shopping circumstance, retailers correspondingly changed their multi-channel marketing strategies to omni-channel marketing strategies in order to meet consumer demand.

Surprisingly, despite the growing important in retail industry, the omni-channel has not been previously conceptualized clearly. Some marketers and researchers have considered multichannel and omni-channel interchangeably and defined omni-channel as one of the multi-channel approaches. However, further review of the omni-channel strategies reveals that omni-channel distinguishes the multichannel in different perspectives (Goulart, 2014).

First of all, omni-channel is not only the broadening the scope of various channels, but also emphasizes on channel integration. Traditional multichannel strategies separate channels apart and consider each channel being independent with other channels and merely sharing customer or inventory information across channels. In the omni-channel world, online channel and offline channels are integrated through product categories, customer information, inventory levels and other information. For example, if a customer picks up a dress shirt that
is out of stock in certain size, the customer service representative in store would be able to immediately check the product availability in the nearest bricks and mortar stores as well online stores. Then the customer could choose whether to get the product from nearest stores or purchase the product from online stores. The product could then shipped to the consumers’ house address or shipped to the selected locations waiting for picking up by the consumer. Once this consumer would neither like to go to another store nor wait for the product shipped from online stores. The representative then could recommend the similar products in the current store based on the customer’s online purchased products as well as shopping cart information.

Second, under the multichannel marketing strategies, the channel performance is evaluated by the sales per channel or customer experience per channel. When it comes to omni-channel marketing, the channel performance is measured by total sales over the channels as well as overall consumer experience over channels (Verhoef, Kannan & Inman, 2015).

Omni-channel strategy is defined by the author as the customer-centric retail strategies that integrating numerous available channels and consumer touch points so as to deliver a seamless, consistent consumer shopping experience, where these channels include brick-and-mortar stores, online stores, mobile stores, mobile app stores, and telephone sales etc. (Figure 1). Survey from an Internet study indicates that customers who shop both online and in stores spend 3.5 times as much as customers who only shop in stores. And consumers who shop through omni-channels (online, in store, and mobiles etc.) spend six times as much as consumers who only shop in stores (McCauley, 2013).
2.4 Consumer decision-making process and touch points

Consumer decision making was defined as “the process of making purchase decision based on cognitive and emotional influences such as impulse, family, friends, advertisers, role models, moods, and situation that influences a purchase” (Schiffman and Kanuk, 2010).

Blackwell, Miniard and Engel (1973) developed a decision-making process model, which consisted of five steps: need recognition/problem awareness, information search, evaluation of alternatives, purchase, and post-purchase evaluation (Hake, 2012).

Then, the consumer decision making model with three stages (Figure 2) – input, process, and output was suggested by Schiffman L., Kanuk L., (2004).
Figure 2. Consumer decision making model (Schiffman & Kanuk, 2004)

The input stage is the external influencers of the consumer’s decision making, which consists of two perspectives: the company's marketing efforts and the sociological environment. The companies' marketing efforts include product, promotion, price, and channels of distribution, while sociocultural environment include family, friends, informal sources, other non-commercial sources, social class, as well as subculture and culture (Schiffman & Kanuk, 2010).
The process stage of the mode explains how customer decisions are made. The psychological factors are the internal factors that influence the decision making process, which consist of motivation, perception, learning, personality, and attitudes.

Under various psychological circumstance, consumers react differently on the external inputs and the behavior differently during the following processes: Need recognition, pre-purchase search, evaluation of alternatives. There are generally two types of need recognitions. One is the actual state type, where products or service fail to meet consumer’s actual need. Another type of need is desired state type, which is consumer’s aspiration to get a new products or service. Many factors can increase consumer’s pre-purchase search such as product factors (frequent changes style and price, many alternative brands, long lasting products), situational factors (first time purchase, gift purchase, social products, and value-related considerations), and consumer factors (demographic and geographic, and personality) (Schiffman & Kanuk, 2010).

The output stage generally contains two activities: purchase and post-purchase evaluation. Consumer information processing strategies could be grouped into two categories: compensatory decision rules and non-compensatory decision rules. Under compensatory decision rules, the consumer will balance positive product attributes with negative products attributes, and come up with an overall evaluation and purchase decisions. In contrast, during a non-compensatory decision mode, certain cutoff point is set to be criteria. Conjunctive rule, disjunctive rule, lexicographic rule are three major single rules used by consumer during a non-compensatory decision process. However, consumers usually involve various decision
roles during a decision process such as conjunctive-compensatory role. Finally, the experience gained from decision making as well as post-decision behavior will in turn modify consumer’s current psychological status. The experience gained through evaluation of alternatives, in turn, affects the consumer’s existing psychological attributes. These consumers may make repeat purchase based on the overall perceived product and experience rating, which is defined as affect referral decision role [Schiffman & Kanuk, 2010].

Traditional consumer decision model has been described as a funnel for more than a century since Elias St. Elmo Lewis, the founder of the Association of National Advertisers built it in 1898. In his model, consumer journey consisted of four steps: Attention, Interest, Desire, and Action (later known as: Awareness & Interest, Search & Evaluation, Purchase, and Post-purchase).

Consumers start at the wide end of the funnel with many brands and gradually narrow them down to a final purchase through brands research and evaluation. So companies used paid-media strategies such as catchy ad campaigns to push the consumer through the funnel, and tried to maximize conversion rate and avoid consumer churn along the funnel.

Until recently, McKinsey challenged the model and pointed out that the model failed to meet today’s digital customer journey. A new customer decision journey model was proposed. This model replaces the funnel with a feedback loop, where customers kept evaluating and experiencing after purchase (C. Edelman & Singer, 2015).
P. Marsden (2011) also proposed a dynamic cyclic consumer decision journey model Figure 3, where the consumer journey is nonlinear, real-time process rather than traditional linear process. Through fast development of digital technology, customers have access to more detailed, efficient information from independent organizations, friends and family. Now consumers consider more brands and dynamically adding and removing alternatives along the journey. Consumers rely more and more on what other brand user reviews – word of mouth and e-word of mouth.

Figure 3. Dynamic Consumer Decision Journey (Marsden, 2011)

A study from McKinsey reveals that a brands’ marketing budgets allocations mismatch with the touch points where consumers are best influenced. The results show 70% to 90% of
marketing spends are used in advertising and retail promotions, which touch consumers’ ‘considering’ and ‘buy’ stages. However, consumers are often more open to be influenced during ‘evaluation’ and ‘enjoy-advocate-bond’ stages. Gallup Poll survey also indicates that 72% of consumers say they trust online reviews. The most powerful impetus to buy is someone else’s advocacy. Thus, a bad review of a brand can totally ruin the consumer’s perceived brand image; the brand may even lose this consumer permanently despite those cool ads (Edelman, 2010).

McKinsey recommended focusing less on spending across media to build awareness and shifting more attention to provide a continuous customer experience and even shortening the journey by eliminating the research and evaluation steps. The final goal is to lock the consumers in the “loyalty loop” (Marsden, 2010).

2.5 Customer experience

Holbrook and Hirschman in 1982 firstly introduced customer experience as a new marketing factor to consider in marketing, and considered the impact of consumer experience such as pleasure and emotions on consumer behavior.

The customer experience is defined as the interaction between a customer and an organization, which is the blend of the organizations’ physical performance and consumers’ sense and emotions across all moments of contacts (Shaw & Ivens, 2002).

Customer experience is viewed as a process by Schmitt (2003), which originates from a set of interactions between a customer and an organization, and then these interactions trigger a
series of reactions. These reactions are very personal that are built up at different levels including rational level, emotional level, sensorial level, physical level and spiritual level (Schmitt, 1999).

The evaluations of consumer experience are based on the comparison between consumer expectations and the actual direct and indirect experiences coming from customer-organization interactions (Meyer & Schwager, 2007, Shaw and Ivens, 2005). The direct experiences happen when consumers buy, experience, and evaluate service from an organization. While indirect experiences are generated during unexpected interactions with companies’ marketing touch points such as sales representatives, recommendations, or resonance with the advertisements.

2.6 Consumer experience measurements

Babin (1994) describes several dimensions of consumer experience dimensions including elements of recreation, enjoyments, and hedonic behaviors overwhelming utilitarian values. Utilitarian and hedonic values represent two dimensions that are used by previous researchers. The utilitarian values measure the consumers’ goal-directed achievements in terms of their investments in time and efforts. While hedonic dimension reflects the entertainment and enjoyment values that the consumers derived from the shopping process. Consumers can thus be divided into two groups: utilitarian consumers and hedonic consumers. Utilitarian consumes are problem solvers that pursuit conscious intended goals, and hedonic consumers are fun seekers emphasize the emotional reward during shopping process. Thus hedonic consumers are more motivated by the process rather than by shopping goals. Later,
researchers (Hoffman and Novak, 1996) examine consumer experience in the online shopping process and claim that hedonic value is more important for many consumers than final utilitarian outcomes in terms of their online behavior.

Schmitt (1999) contrasts traditional marketing methods and proposes a strategic model called experiential marketing. Experiential marketers consider consumers as rational and emotional human beings who are concerned with experiences. There are five types of experiences for consumers: sense, feel, think, act and relate. Sense module refers to the sensory experience created through sight, sound, touch, taste and smell. Feel module stands for consumers’ inner feelings and emotions, and marketers would like to create effective experiences. Think marketing aims to create cognitive, problem-solving experiences through engaging customers creatively by surprise, intrigue and provocation. Also, the relate module includes aspects of sense, feel, think and act modules, and refers to the individual’s desire for self-improvement.

Gentile et al. (2007) confirm the experiential marketing model and states that customer experiences are created by hedonic and utilitarian consumer values, as well as the companies’ own value. There are six modules in their model: sensorial, emotional, cognitive, pragmatic, lifestyle, relational module.

Later, Verhoef et al. (2009) develop a customer experience creation concept model and propose several dimensions of customer experience including: social environment, service interface, atmosphere, price, assortment and channels. Also, past customer experiences have an impact on current and future customer experiences.
Kim et al. in 2011 develop the customer experience index (CEI) that is based on service delivery system in general industry. The dimensions in CEI model are: environment, benefits, convenience, accessibility, utility, and incentive.

However, through Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), Klaus & Maklan’s (2012) customer service experience (EXQ) model removes social environment, retail atmosphere and retail brand factors in Verhoef’s (2009) model and four dimensions are included in EXQ model: product experience, outcome focus, moments of truth, and peace of mind.

Previously, most researches on customer experience are conducted from the perspective of the overall evaluations or values based on the accumulation level. This study perspective restricts the understanding of the insights between the consumer and the retailer. Stein and Ramaseshan (2015) identify, categorize, and define the dimensions of consumer experience from consumer’s touch points perspective. As discussed in the consumer decision making process section, consumers shop through a decision making journey and retailers are engaging with the consumer constantly at different touch points in the journey. And from the consumers’ perspective, the touch points represent what actually happens during their shopping journey on different channels. Based on a qualitative research method, Stein and Ramaseshan (2016) uncover seven dimensions of consumer experience: atmospheric, technological, communicative, process, employee–customer interaction, customer–customer interaction, and product interaction elements.
This study extends Stein and Ramaseshan’s (2016) study, and proposes a validated touch-point based customer experience measurement model (TPE) for omni-channel consumers.

Touch point based experience (TPE) model (Figure 4) consists of seven modules in Stein and Ramaseshan’s (2016) qualitative study. Specific definitions and corresponding objectives are discussed as follows.

1. Atmospheric touch point

Atmospheric touch point is defined as the physical characteristic and surroundings consumer may observe when interacting with the retailer. There are various atmospheric elements at different touch points, at different stage of the decision making journey, and on different channels. These elements include store amenities, store attractiveness, store layout and design, and store display (Ramaseshan’s, 2016).

Store atmosphere and its impacts have been widely studied for decades, and store atmosphere was found to have significant effects on sales (Milliman, 1982; Stanley & Sewall, 1976). Also, store atmosphere is positively related with customer satisfactions (Bitner, 1990). Store environment elements such as layout, color, lighting, and music are also significantly impact consumer intention, satisfaction, consumption amount, and perceived quality (Babin & Attaway, 2000)

Research reveals that customer perceptions of store environment such as flooring and lighting significantly influence customers’ shopping experience (Terblanche & Boshoff’s, 2001; Bagdare, 2014)
From omni-channel consumer’s’ point of view, atmosphere not only refer to store environment, but also refer to online, mobile atmosphere and Brakus et al.’s (2009) study implies that website content and design could impact consumer’s brand experience.

Thus, an aggregated atmospheric touch points from all shopping channels include in-store, online, mobile atmosphere, are expected to influence omni-channel consumer’s shopping experience. Driven by this assumption, we introduce the following objective:

**RO1: To investigate the influence of atmospheric touch point on omni-channel consumer experience.**

2. Communicative touch point

Communicative touch point is communication from the retailer to the customer, which includes promotional and informative messages (Ramaseshan’s, 2016). This information includes price promotions, special offers, and personalization. Promotional and informative messages were found to directly influence customer satisfaction, which in turn influence brand loyalty (Zhang and Tang, 2010).

Zakaria et al (2014) examine various loyalty programs and find that there are positive and significant relationship among loyalty program, brand loyalty and brand satisfaction.

Bojei et al. (2013) study the marketing tools and their impact, the result show that loyalty/rewards programs and personalization have a significant and positive relationship with customer retention.
We assume that communicative touch point in omni-channel retailing era will continue influence consumers behavior. Thus we have the following objective:

**RO2:** To investigate the influence of communicative touch point on omni-channel consumer experience.

3. Product touch point

Product touch point refers to the direct and indirect interactions consumer have with the core tangible and intangible products, which include product quality and product assortment etc. (Ramaseshan’s, 2016). Product assortment refers to retailers’ ability to provide a wide range and product choices (Verhoef et al., 2009). Product assortment is found to have a positive impact on customer experience. Research also reveals that quality is the key factor to drive customer satisfaction (Gomez et al., 2004). Thus, we expect that product touch point will positively influence omni-channel customer experience and introduce the following objective:

**RO3:** To investigate the influence of product touch point on omni-channel consumer experience.

4. Technological touch point

Technological touch point is defined as a customer’s direct interaction with any form of technology during the shopping journey (Ramaseshan’s, 2016). Technology based marketing is playing more and more role in retailing industry online, on mobile devices and in bricks and mortar stores. And retailers are increasingly implement technologies to improve
productivity and service quality. Weijters et al. (2007) take four factors when measuring self-service technology: perceived usefulness, perceive ease of use, reliability and fun associated with the technology. Their study reveals that all these four factors are positively related with customer attitude toward the technology, which in turn affects customer satisfaction.

Hsu (2008) found that the quality of technology based service has significantly positive influence on customer relational benefits such as brand loyalty.

Based on previous literature, we expect that technology based touch points will continue impacting omni-channel consumers shopping journey. Thus the following objective is introduced:

**RO1:** To investigate the influence of technological touch point on omni-channel consumer experience.

5. Process touch point

Process touch point stands for the actions or steps consumer need to take during shopping journey with a retailer. Process elements include waiting time, navigation and service process etc. (Ramaseshan’s, 2016). Weijters et al. (2007) found that perceived waiting time is an important antecedent of customer satisfaction in the shopping journey. In terms of online shopping, a company’s website is both a communicative and a sale channel, Shim et al. (2015) found that favorable online shopping flow positively influence sensory and affective brand experience, which in turn led to brand loyalty. For mobile shopping process, Zhou et al. (2010) found that information quality and system quality significantly affect users’ trust
and experience, which future influence brand loyalty. Kim and Yu (2016) also found that the involvement level of branded applications on mobiles has a significant effect on the relationship between brand experience and brand loyalty. Thus the positive relationship between process touch points in omni-channels and customer experience is expected. We then investigate that:

**RO5**: To investigate the influence of process touch point on omni-channel consumer experience.

6. Employee-customer touch point

Employee-customer touch point is the direct or indirect interactions customers have with retailers’ employees (Ramaseshan’s, 2016). Previous studies have found that customer perception of employee interpersonal behaviors such as listening behavior, customer orientation could affect customer satisfaction (Dean, 2007; Dagger et al., 2007). Interpersonal bonds are also strong predictors of loyalty in service (Gremler and Brown, 1996). Customer perceptions of employee performance are found positively influence customer satisfaction and loyalty. Rabbanee et al. (2014) find that employee performance positively affects customers’ perceived value, trust and loyalty.

Based on the previous literature, we assume that employee-customer touch point will positively influence omni-channel customer experience to some extent. Thus we have the following objective:
RO6: To investigate the influence of employee-customer touch point on omni-channel consumer experience.

7. Customer-customer touch point

Customer-customer touch point is the direct and indirect interaction between the consumer and with other consumers (Ramaseshan’s, 2016). There are different kinds of customer-to-customer interactions such as word-of-mouth, customer reviews, and e-word-of-mouth. Researchers have found a positive relationship between customer-to-customer interaction and customer satisfaction (Harris et al., 1997; Arnould and Price, 1993). Also customer-to-customer interaction is positively related to brand loyalty, word of mouth, and customer experience. (Chen et al., 2012; Lloyd and Luk, 2011; Harris and Baron, 2004). In addition, Haytko and Simmers (2009) find that online customer-to-customer interactions significantly impact customer satisfaction. Hence, we expect that in an omni-channel shopping environment, an aggregated level of customer-to-customer interaction is likely to have a positive impact on customer experience. Thus we have the following objective:

RO7: To investigate the influence of customer-customer touch point on omni-channel consumer experience.
2.7 Omni-channel consumer behavior

According to Verhoef (2012), the scope of omni-channel consumer behavior studies lies in three main areas, which are the consumer behavior across channels, the impact of channels on performance, and retail mix across channels.

2.7.1. Consumer behavior across channels

First of all, different consumer behavior patterns emerge across various product categories. A PWC (Price Waterhouse Coopers) study shows that for certain product categories, consumers still prefer to shop via a single channel. For instance, half of consumers prefer to purchase groceries in brick and mortar stores because of the ability to select the product at first hand such as meat or vegetables. For the books, music, and films categories, nearly two third of the respondents prefer to buy online because of the product familiarity and convenience. Electronics and furniture/home goods shoppers tent to leverage different channels before making the final purchase, which may due to the wide product selection, the importance of the user experience and relative high cash outlay. When it comes to the apparel and footwear categories, 71% of the respondents would shop across different channels, either online or in stores (PWC, 2015).

Also, some researchers segment the omni-channel consumers based on their demographic information as well as lifestyles. Crosett (2009) proposed three consumer segmentations that are most likely to be omni-channel shoppers. Young mobile shoppers are people younger than thirty who use text messaging. Omni/integrated shoppers tend to be affluent, home-
oriented, and usually 30-50 years old. Social networkers are the consumers connected by specific topics both on social media as well as in daily life.

2.7.2. Impact of channel choice on retail performance

MacDonald, Wilson, and Konus (2012) challenge traditional quantitative data from survey as well as qualitative methods such as focus groups, and claim that both methods rely too much on consumer’ memories, which may decay real quickly. They introduce an innovative way to track consumers called real-time experience tracking (RET). This system relies on a quick SMS-based four-character micro-survey filled out by the consumer every time they notice a brand. These experiences include attracting by an advertisement, making a transaction, or sharing brand information with other consumers. The result shows that some touch points positively influence consumers purchase intention. For example, consumers who encounter the product in a friend’s house are two more times as likely to purchase it compared to people who did not encounter the product. Their study also suggests that in-store communication have a stronger influence on consumers brand preference than other touch points (Macdonald, Wilson & Konuș, 2012).

Using secondary data in the US retail industry, Cao and Li examine the impact of channel integration on retail performance and report a positive relation between channel integration and retail performance. Other researchers study the influence of mobile usage on consumer purchase behavior and find that mobile channel usage affects consumer behavior across channels, and consumers use mobile device mostly during search phase.
2.7.3. Channel integration on channel performance

Researchers try to shed light on the impact of channel integration on channel performance. Herhausen et al. leverage technology adoption research as well as technology diffusion theory and find that channel integration performance is influenced by consumers perceived service quality, perceived risk, as well as overall online shopping experience. The result also indicates that channel integration leads to synergies of channel performance rather than cannibalization. Other studies test how price discounts in one channel affect sales for the same product in another channel (Gong, Smith and Telang, 2011). They found that price promotions in one channel could increase the brand awareness in another channel, and thus offset the cannibalization effect of another channel.

Verhoef et al. (2015) discuss the study on the impact of the three different assortment strategies on consumers perceived shopping benefits and patronage intentions. The first assortment strategy is no integration of channels, which means different channels offer different products. Asymmetrical integrations, also called long tail integration is another assortment strategy that most retailers have employed, where online channels provide different additional merchandize than products in stores. The last assortment strategy is full integration when online stores offer same product with off-line stores. Their study reveals that for assortments with substitutive relations, full integration performs better than no integration, and asymmetrical integration. However, when the assortment contains a wide range of complementary relations, full integration is not always effective than asymmetrical integration (Verhoef, Kannan & Inman, 2015).
Based on the above discussion, it is reasonable to expect that consumers’ channel choice will have a significant influence on consumers’ behavior as well as the relationship between consumers touch points and consumers experience. Thus we have the following objective:

**RO8: To investigate the influence of consumers’ channel choice (single channel VS multi-channel) on the relationship between touch points and omni-channel consumer experience.**
2.8. Theoretical Framework

Based on the above literature review and objectives, a theoretical framework (Figure 4) was developed as following:

Figure 4: Touch points based experience (TPE) theoretical framework for research objectives 1-8
Part B. Omni-Channel Consumer Engagement Strategies

From the above discussion, there is no doubt that more and more consumers in the omni-channel retailing market are value-sensitive, mobile-powered, channel-unconscious, and social-active. For many years, retailers have relied on marketing data to improve business performance and make decisions. However, retailers nowadays find it difficult to keep competitive advantages solely based on traditional marketing strategies such as push promotions and traditional funnel-based marketing approaches focus on driving consumer awareness and consumer acquisition rather than building a satisfied consumer shopping experience (Sridharan, 2012). The explosion of consumer data has brought up opportunities for retailers to turning data into actionable, accurate, and valuable business insights. The most often used consumer analytics technologies and strategies are demonstrated in the following content.

Leading apparel and footwear retailers are integrating both traditional methodologies and cutting edge data analytics technologies to discover fashion trends, patterns, and consumer insights (Stodder, 2014). Consumer data analytics allow the retailers to understand and predict consumer’s action during their shopping journey. Then based on these insights, the retailers could design consumer-centered programs to drive acquisition, retention, recommendation, as well as targeted marketing. Examples of the consumer analytics applications include customer segmentation, targeted promotion, next-best-offer analysis, etc. (Sridharan & Bateman, 2013).
2.9. Consumer data model

During a consumer’s shopping journey, a large amount of data can be gathered through consumer interaction (Figure 5). The consumer data can be divided into four categories: interaction data, descriptive data, attitudinal data, and behavioral data, which explain where, who, why, and what. Interaction data include e-mail and chat transcripts, call center notes, web click-streams, and in person dialogues transcripts. Descriptive data include geo-graphics, demographics, self-declared information, characteristics, etc. Attitudinal data consists of market research, social media data, WoM (word of mouth) as well as E-WoM. Behavioral data include orders, past transactions, usage histories and so on. In order to deal with such large volume, highly dynamic data, many companies have employed both traditional methods as well as cutting edge data analytics software or service, such as cloud computing, Hadoop, data virtualization, and data visualization (IBM, 2013).
2.10. Consumer marketing strategies review

a. Consumer segmentation

In order to identify what market campaign works most effectively, companies need a detailed understanding of the attributes for each customer segment. According to Gilboa (2009), the best method to study consumer behavior is consumer segmentation, which was introduced by Wendell Smith in 1950s. Consumer segmentation assumes a group of people share similar
shopping habits as well as product preference, and based on these similar characteristics, customer are divided into different groups.

Four factors are commonly used for consumer segmentation, which are geographic, behaviors, demographic, and lifestyle. Geographic factors indicate region of a country, population, location coordinates, and weather. Behavioral segmentation considers variables like: product attributes, purchase history, reward program participation, etc. Demographic information includes age, gender, income, education, etc. Lifestyle factor measures attributes like social values, self-concept, and community affiliations (Megento, 2016).

More and more companies segment consumers based on product-related factors rather than demographic information because for instance, totally different behavior could exist within the same age group (Delft, 2013). Through past purchase history and web click streaming, consumers are shown a range of different products. This segmentation method could increase store traffic and improve overall consumer satisfaction.

b. Recommendation engine

The recommendation engine is the information filtering tool that could predict what a user may prefer among a list of items. As an alternative to search engine, the recommendation engine guides the consumers to discover products or service that may not be noticed during a search process. Recommendations have contributed more and more revenue than previous for giant retailers such as Amazon, and Walmart (Linden, Smith & York, 2003).
One of the most famous examples for recommendation is Amazon’s “People who bought this also bought” engine. Based on user’s items based collaborative filtering (IBCF), Amazon reported twenty percent more revenue via recommendation engine (Hinssen, 2012).

Collaborative filtering and content-based filtering are the two major methods to filter the information and create a list of recommended products (RIDWAN, 2015).

Collaborative filtering methods are based on consumers’ previous purchases, selected or rated items as well as similar behavior shown by other consumers; then a list of products is recommended to the consumer. One of the biggest advantages of collaborative filtering method is the independency of reliance on item’s specific attributes, which is particularly suitable for items that have complex characteristics such as movies.

Content-based filtering approaches compare the similarity between item attributes and then recommend products that are close to the current or past selected product. In addition, a user profile is created to indicate the type of products this consumer likes. During a content-based filtering process, various products are compared with the user’s preferred products in the profile and the best matched items are recommended.

However, due to some common problems in recommendation systems such as cold start and the sparsity issue (particularly for new products), more and more inaccuracy recommendations occurred for a single recommendation approach. Researchers found that a hybrid recommendation method, which combines collaborative filtering and content-based filtering approach, could provide more accurate recommendation results. The combination
could be either firstly separating running each algorithms then combining the results or implanting one method on the other method. The online streaming movie retailer Netflix is current employing this hybrid recommendation algorithm, where they compare the users that sharing similar watching or search habits as well as recommend movies that have similar attributes with film that the consumer has previously rated with high scores (Dataaspirant, 2015).

c. Cross-selling, up-selling

Cross-selling generally occurs when more than one type of products are offered to consumers in order to satisfy additional, complementary needs that are unfulfilled by the original item. For example, a customer who buys a shirt may be offered ties by the sales representative with a certain amount of discount.

Up-selling is one of the sales strategies that consumers are offered a higher-end version of the product than the current selected products. For instance, the sales representative may introduce durable wind, water, solid, and oil repellent outdoor jackets rather than common wind-proof ones that the consumer may initially trying to purchase.

Traditional retailers have been cross-selling and up-selling for many years with the recommendation of sales representatives. During the cross-selling, the products are recommended based on the consumer’s pre-defined or potential defined need, and how the recommended products could fulfill this need. Up-selling strategies are mainly about how to build values to the recommended products (Lewis & Media, n.d.).
In ecommerce, cross-selling is often presented as “Frequently bought together” during the check-out stage. Cross-selling and up-selling can remind consumers of the breadth and depth of a catalog, drive store traffic, and increase repeat purchase. Cutting edge cross-selling methods utilize predictive analytics to determine which consumer or consumer segmentations most likely buy what cross-selling or up-selling product.

A study from Forrester indicates that products recommendations like up-sells and cross-sells account for approximately 10-30 percent of ecommerce revenues (Bigcommerce, 2015).

The latest trend of cross-selling and up-selling is in-store recommendations with mobile devices powered by pre-set algorithms. Through the hand-held devices, the sales representatives could recognize the consumers as well as their purchase preference, and then recommend the automatically generated products. Also, the staff could also look up the stock level and delivery schedules, so that the cross-selling and up-selling could be accomplished in real time through both in-store as well as online. Survey shows that nearly sixty percent of retail consumers are not willing to queue for service, among which about thirty two percent consider leaving and buy online, and eighteen percent will buy at another store. These hand-held mobile tools can also help the sales assistant check out in real-time, eliminate the queue time, which in turn increase consumer satisfaction. Argos is one of the retailers who are taking the first step. Utilizing iPad-based kiosks and empowering staff with iPad, Argos help their consumer choose and select the product. By examining the products associations, sales representatives can determine which joint offers are more likely to be purchased, then provide some discounts to stimuli the transaction. At the same time, consumers who buy
products online will be informed of the nearest stores where they can test and try the products. These processes greatly simplify and speed up the sales process (Cook, 2014).

However, Denish Shah and V. Kumar’s study revealed the dark side of cross-sell and they proposed four distinct customer profiles that could cause the loss of revenue: service demanders, revenue reversers, promotion maximize consumers, and spending limiter. Further, they suggest that before undertaking cross-selling or up-selling initiatives, companies need to analyze the past transaction data to determine whether the customer or customer segmentation belongs to unprofitable consumer profiles (Shah & Kumar, 2012).

d. Predictive analytics- targeted promotions, churn prediction, planning optimization

Predictive analytics is a data mining method that uses analytical models to discover hidden patterns and apply these patterns to predictive future trends and behaviors, then decisions and strategies are made based on these analysis (Quickborn, 2012).

Retailers are now gathering both structured data such as transaction data as well as unstructured data like social media data. Predictive analytics help these retailers understand the deeper insights of consumers and make more efficient market campaigns. Survey shows that retailers who make decisions based on predictive analytics achieve significant higher return on investment and retain more loyal consumers (Vesset, 2011). Three important applications of predictive analytics are discussed as follows.

The first application of predictive analytics is targeted promotions. By capturing consumer’s purchasing history data, web-clicking data, as well as social media interaction data, the
retailers can segment the consumer into groups or even individuals, and provide personalized promotions or offers to this consumer. For example, targeted promotions of prom dresses can be sent to consumers who are going to graduate.

Second, predictive analytics enable retailers to analyze consumer behavior patterns and find out which consumers are most likely to leave for competing retailers. Then the algorithms will calculate whether actions should be taken to keep the consumer. For instance, a mobile carrier from Europe tracking a sudden surge of activities on competitor’s website, and identify consumers who may leave for this event. Then the mobile operator immediately launches counter promotions that successfully prevent consumer churn (Hinssen, 2012).

Last, predictive analytics help determining the optimized stock levels at each online and offline store. Based on the comprehensive factors such as demand forecasting, current SKU levels, market baskets, as well as store location, the retailers can optimize the allocation of the inventory, even balance the inventory by store-to-store shipping to prevent out of stock and overstock. At the same time, clustering analytics group similar stores together to optimize stock levels, increase inventory turnover and enhance customer experiences (Linksy, 2012).

The Scenario presented is only part of the benefits that the retailer could enjoy by deploying predictive analytics into a retailer’s operations and systems. The following lists out other areas where competitive advantage can be gained:
1. Retailers could run market basket analysis, churn analysis and mine consumer loyalty data to figure out key performance indicators that drive customer retention and campaign optimization (Siegel, 2008).

2. Retailers may also sense and shape consumer demand by optimizing price and inventories. Also, the next best actions could be mathematically calculated in a specific situation based on market campaign goals and constraints. This campaign may lie in a sweet spot so that maximized profits could be achieved while not overexposing the retailers to risks. Through the next best actions, retailers could use just forty percent mailing to achieve eight percent of responses (FICO, 2010).

3. Personalized coupons can be generated at the point of sale (POS), based on the customer’s current and past purchases, to increase redemption rate.

4. Predictive analytics can determine root causes of failures issues, such as loss profits during a market campaign, and improve the quality of risk management (Chase, 2013).

e. Sentiment analysis

Sentiment Analysis, also known as opinion mining, is the process of extracting, identifying, or characterizing the text unit, in order to determining whether the attitude is positive, negative or neutral (Lexalytics, 2016).

As online and off-line consumers are increasingly discussing their purchases on social media platforms, the ability to analyze consumers’ attitude has become crucial.
Using natural language algorithms and web data extraction tools, retailers now can follow conversation on social media as well as analyze comments on retailers’ sites, which allow the retailers to understand the feeling in consumer’s mind about the products, brand, or service. Also, the sentiment analysis could be combined with point-of-sale data to better understand how e-WoM (electronic word of mouth) and consumer’s attitude influence their behavior during the shopping purchase decision process. Retailer generated explanation and complementary comments can correspondingly be added to the website to eliminate consumers doubts (Hinssen, 2012).

However, comments on social media and retailer’s websites filled with colloquialisms, misspellings, Internet slang and sarcasm, which makes it inaccurate for traditional text analytics tools to determine the attitude of the texts. State-of-art Natural Language Processing (NLP) engines are built and trained to deal with complex sentiments. When integrated with text analytics and machine learning algorithms, NLP shows significant better accuracy than traditional approaches (Netbase, 2013).

f. Location based marketing

As mentioned previously, three trends have changed consumers shopping habits and the way retailers interact with their consumers. First, consumers routinely interact with brands through different channels using various devices. Second, consumers no longer merely rely on the retailer’s websites; they consult more from third party website and from social media. Last, as the result of the popularity of smartphones and more precise GPS-based service, consumers rely more on mobile-location based search engine during shopping process.
SoLoMo, short for social-local-mobile, is a new mobile-location based marketing approach that allows the retailers to push notifications to consumers who are geographically nearby (Techopedia, n.d.). SoLoMo applications allow advertisers to push notifications to potential customers who are geographically nearby. Groupon and Yelp are examples of SoLoMo (Rouse, 2013).

Fashion retailers such as Neiman Marcus' have implemented SoLoMo App (social, local and mobile) in bricks and mortar stores to help creating a unique “co-shopping” experience. This mobile application allows the retailers to track consumers in stores, and helps the retailers to inform consumers about where is the sales associates, store events, promotions, new products, etc. In addition, consumers can also select products, check the inventory and scan QR codes. The sales associates could then use the information generated during this process to offer more personalized promotions (Reddy, 2015).

Retailers even could send personalized information while consumers are near the local stores. For instance, GPS-integrate beacon system can first detect that the consumers are in close proximity to stores. Then a personalized offer could send to consumers who are opts-in to receive marketing notifications. As interested consumers head to stores and purchase the products using the promotion codes, another cross-selling offer will then be send to the sales representative or directly to the consumer (IBM, 2011).

Other leading-edge retailers install cameras systems in stores, and analyze video streams generated from these systems, then draw the mappings of the consumer’s foot path. This
video data analysis is then developed into applications that help the retailers to optimize product placement in stores (Hinssen, 2012).

8. In-store digitalization

Survey shows that most customers still preferring buying apparel and footwear in-stores, however, the long check-out queue is a pain spot for consumers. Retailers are increasingly leverage digital technology to intervene along consumers journey in order to deliver a seamless omni-channel shopping experiences. NFC (near-field communication), digital wallet, mobile POS, virtual try-on room, digital kiosks are state of art technologies employed by different retailers. Several examples of digitalized in-store retailers are discussed below (Anand, 2015).

Nordstrom, as one of the largest fashion retailers has spent more than thirty percent of its capital budget on technology. The “Nordstrom Innovation Lab” tracks Pinterest pins to capture the latest trends and promote relevant products through social media to drive online and in-store traffic. Also, Nordstrom merges in-store experience with social media by allowing consumers to see pictures and products reviews on Instagram through a large in-store screen. “Co-shopping” experience is the concept pursuit by the retailer, where sales staff and consumers are “shopping together”. Employing new technologies such as sensors and Wi-Fi signal, the retailer could track the consumer’s behavior, and sales representatives are helping the consumers based on these analysis. Also, Nordstrom integrates online inventory with off-line inventories to help keeping a unified shopping experience.
The luxury retailer, Rebecca Minkoff, embedded touch screens in store mirrors, where consumers could select the products in the screen and the product could be sent directly to fitting room. Using RFID technology, sensors and interactive displays, this retailer takes the in-store inventory management to a new level.

SportCheck is Canada's largest sports retailer, who is known for digitally-enhanced stores. For example, In Toronto’s digital lab store, there are seven hundred digital screens built everywhere displaying different products ranges, special personalized offers, as well as local sports events. Once a customer steps onto specialized sensor, the monitor will show recommended shoe sizes and recommend special offers. Also, digital kiosks allow consumers to customize their shoes as well as sunglasses (Reddy, 2015).

Through employing new technologies and launching market campaigns, fashion retailers are continuously improving their omni-channel consumer experiences. For example, partnering with Postmates, American Apparel became the first retailers to offer 1-hour delivery. Consumers are able to receive apparel and footwear within sixty minutes’ delivery window. 79 American Apparel stores throughout 31 metropolitan areas are participating in the services. American Apparel’s next phase is to integrate the omni-channel experience with RFID technology for real time inventory check (Dickey, 2016).
h. Tools and algorithms

Algorithms are the engines of the market analytics tools. Survey conducted by Rexer Analytics suggest that the most used algorithms by companies during data analysis are regression, decision trees, cluster analysis, time-series, neural networks, associations rules, support vector machines, and Bayesian (Rexeranalytics, 2010).

During the past decades, critical efforts have been made to optimize the sales promotions or recommendations. As discussed above, the methods used by researchers have included classical statistical methods such as Bayesian probability models, machine learning and data mining techniques such as k-NN, ANN, SVM (Johnson, Tellis & Ip, 2013). Close examination of the promotional efficiency of these machine learning models reveals that for different product categories, e.g., milk and beer, these algorithms show different reliability and validity (Soguero-Ruiz, Gimeno-Blanes, Mora-Jiménez, Martínez-Ruiz & Rojo-Álvarez, 2012).

An omni-channel graphic model (Figure 6) is built based on the reviews of current omni-channel consumer interaction strategies as well as the algorithms behind these strategies. This graphic model illustrates how retailers can engage with the omni-channel consumers during their purchase decisions.
Figure 6. Omni-channel consumers’ engagement strategies review graphic model

i. CLV (customer lifetime value) and RFM (recency, frequency, monetary)

One of the rules to segment consumers and make market strategies is based on customer lifetime value, which is defined as the net present value of cash flows contributed by a consumer or consumer segmentation over their life of relationship with an organization.

Researchers have pointed out that customer lifetime value can be set as an accurate guide for
consumer acquisition and retention. Also, consumer segmentation can be conducted based on
different level of consumer lifetime value (Kristiani, Sumarwan, Noor Yuliati & Saefuddin,
2013).

Customer lifetime value (CLV) has previously been calculated in multiple methods at an
aggregate level. The fundamental approach to calculate customer lifetime value is illustrated
in Equation 1. Further, research has shown that if margins \((p - c)\) and retention rates are
constant over time, and the time is infinite, then the CLV expression could be simplified to
Equation 2. This equation indicates CLV equals to margin multiply by a constant (Gupta, etc.,
2006).

\[
CLV = \sum_{t=0}^{T} \frac{(p_t - c_t)r_t}{(1 + i)^t} - AC
\]

(1)

where:

\(p_t\) = price paid by a consumer at time \(t\),

\(c_t\) = direct cost of servicing the consumer at time \(t\),

\(i\) = discount rate

\(r_t\) = probability of consumer repeating buying at time \(t\)

\(AC\) = acquisition cost

\(T\) = time span for estimating CLV
\[ CLV = \sum_{t=0}^{\infty} \frac{(p - c)r^t}{(1 + i)^t} = m \frac{r}{(1 + i - r)} \]  

where:

\( p \) = price paid by a consumer at time \( t \),

\( c \) = direct cost of servicing the consumer at time \( t \),

\( i \) = discount rate

\( r \) = probability of consumer repeating buying at time \( t \)

\( m \) = margin at time \( t \)

Gupta et al. (2006) summarize six modeling approaches that measure customer lifetime value. The first approach is probability models, which is based on Pareto/NBD model and Markov chains. The second method is based on Pareto/NBD model and customer acquisition, customer retention, and customer margin and expansion. These kinds of model are based on econometric models like probability models. Third, by modeling the marketing components such as consumer acquisition, retention and cross-selling, a persistence model is developed. The fourth modeling method is to use aggregate data and diffusion or growth models to predict the number of consumers that the companies would acquire. Another method is based on computer science models such as data mining, machine learning models. The last method
is probably the mostly frequently used approach to calculate CLV, which is a RFM (Recency, Frequency, Monetary) based calculation (Gupta, etc., 2006).

Bult and Wansbeek (1995) defined RFM as: R (Recency): the time period since the consumer’s last purchase. A lower value shows a higher probability of the customers to make repeat purchase; F (Frequency): total number of purchases made within a certain time period. A higher frequency means higher consumer loyalty toward the brand; M (Monetary): the money spent during a certain time period; A higher value indicates more valuable consumers.

Previous researches claim that RFM is one of the most accurate methods to calculate CLV and predict consumers’ future behavior since it measures when, how often, and how much they purchase a brand. RFM method begins with sorting and ranking customers from lowest to highest based on Recency, then splits the customer into quintiles, then assigns value to the lowest scored quintile to highest scored quintile from 5 to 1. When it comes to Frequency and Monetary, the lowest scored quintile becomes 1 and highest scored quintile becomes 5. RFM analysis could segment the customers into maximally 125 (5×5×5) groups. The best customer group is 555 group, with most recent purchase, most frequently, and most money spend. Thus, targeted promotions could be sent to corresponding groups to retain a long-term relationship.

Recent studies proposed WRFM –Weighted RFM- instead of RFM as more accurate method to calculate CLV for marketing strategies. On one hand, some researchers suggest listing Frequency as the top priority and putting less credit (weights) on Monetary. On the other hand, Monetary factor is viewed as the most important attribute in Chuan and Shen’s (2008)
study. In a WRFM calculation, different weights are assigned to RFM parameters based on the industries (Equation 3) (Kristiani, Sumarwan, Noor Yuliati & Saefuddin, 2013). Analytical hierarchy process (AHP) method is usually used to set the specific value of weigh of RFM (Khajvand, Zolfaghar, Ashoori & Alizadeh, 2011).

\[ Y = a \times (R) + b \times (F) + c \times (M) \]  \hspace{1cm} (3)

where:

\[ a = \text{the weight of Recency,} \]
\[ b = \text{the weight of Frequency} \]
\[ c = \text{the weight of Monetary} \]
\[ R = \text{Recency} \]
\[ F = \text{Frequency} \]
\[ M = \text{Monetary} \]

The Analytic Hierarchy Process (AHP) was first developed by Thomas T. Saaty, who is a current professor at University of Pittsburgh and former professor at the Wharton School, University of Pennsylvania. The analytic hierarchy process is a multi-criteria decision-making (MCDM) method, which derives priority scales through pairwise comparisons and
relies on the judgments of experts. It allows decision makers to model and solve a complex problem in a hierarchical structure (Wang & Zhang, 2013). According to Saaty (2008), in order to generate priorities, decision makes should decompose the decision in to four steps: a) Define the decision problems and determine the knowledge scope b) Structure the problems in a hierarchical way, with the goal on top followed by objectives, then criteria and sub-criteria, and finally with alternatives at the lowest level c) Construct a set of pairwise comparison matrices d) Weight the priorities from top objectives to the final priorities of the alternatives.

However, there are several drawbacks of AHP model, which make the model less reliable when there are large number of dimensions. First of all, AHP model assumes that the relationships between clusters of different levels are unidirectional and there are uncorrelated relationships between clusters as well elements within one cluster. In the real world, interdependence can occur in many ways. Second, the maximum number of criteria range between 5 and 9 max (bpmsg, 2013). Larger sets of pairwise comparison will cause problems of a) time consuming due to the difficulty to build pairwise comparison consensus b) consistency problem due to human judgement variations c) small changes in comparison lead to big changes in ranking (Opydo, 2014). However, real life scenarios often expose us to very large number of criteria. Thus AHP model is not an efficient and reliable way when dealing with large dimensions’ datasets. Further, even though there are increasing number of researchers including RFM variables in developing clustering and classification models, RFM weights are mostly calculated at an aggregate level, for example, RFM weights for a
whole industry or RFM weights based on year unit (Olson, Cao, Gu, & Lee, 2009). It is a trend for companies to find a balanced spot of RFM weight and implement it to individual level, rather than aggregate level, whose goal is to build marketing strategies at individual level (Curtin & Ogden, 2012).

This study will explore a new method to generate RFM weights, compare RFM weights and calculate customers’ purchase probabilities at individual level, so that companies could send targeted promotions in more niche ways. We thus have the following research objective:

RO$_9$: To investigate the role of RFM on predicting omni-channel consumers’ purchase probability for targeted promotions.
CHAPTER 3
Research Methodology
This chapter presents the methodological procedure utilized to collect data and the instruments used to measure the variables. In addition, measurement tools and methods used in the research are discussed.

3.1 Data collection

Because of the broad scope of the topic, two datasets were collected in the research. The first data set is primary data, which was collected through the form of a questionnaire. The second data set is secondary data that was obtained through a proposal to The Wharton School of the University of Pennsylvania. The data collecting procedures are explained in this next section.

3.1.1 Research Objectives 1-8 data collection: Survey questionnaire

Survey questionnaires were used to collect the primary data to investigate omni-channel consumers insights and their expectations of retailers market strategies. Before the survey was administered, the Institutional Review Board of North Carolina State University approved the survey procedures including reaching the population and the form of the questions (Appendix A).

The survey (Appendix B) was then created on an online survey platform called Qualtrics (https://www.qualtrics.com). Then a QR code was generated that could direct the participants to the online survey. There were two ways to distribute the QR code, either the
researcher randomly asked the students at an eastern university in U.S.A. whether they would be willing to fill out a questionnaire through scanning the QR code or an Email link contained the QR code was sent to students from this university. One concern is that student samples limit the generalizability and applicability to the general public consumers, but college students represent typical omni-channel consumers who are relevant in the present study. All questions were measured by a five point Likert scale. The risks of participating in this study were no greater than those encountered in daily life. All the participants were kept anonymous and no identifiers could be used to identify the participants.

3.1.2 Research Objective 9 data collection: The Wharton School dataset

The secondary dataset was obtained by submitting a proposal (Appendix D: Wharton Data Proposal Submit Link) to The Wharton School of the University of Pennsylvania. The Wharton Customer Analytics Initiative (WCAI) is an academic research center that focuses on better understanding customers through analyzing the individual-level data collected by companies. The proposal was submitted to a Fortune 500 Specialty Retailer’s project, who tries to identify consumers’ daily life projects. The retailer believes that when consumers are in the midst of a project, they are more open to product suggestions and recommendations to make the projects more successful. The data set contains 10 million item level purchased history transactions data from 60000 consumers over two years from March 2012 to March 2014. In addition, detailed product hierarchies and product attributes, store locations, and email campaigns information were available in the dataset (WCAI, 2014).
3.2 Research Objectives 1-8 survey Instruments

There were two sections in the questionnaire. The first section measured the consumers’ touch points include atmospheric touch point, communicative touch point, product touch point, technological touch point, process touch point, employee-customer touch point, customer-customer touch point. In addition, general omni-channel customer experience was measured in the survey. Existing scales from previous research were adopted or modified to measure the constructs. The second section included the demographics information such as gender, age, education. Below is the detailed information of the measurement of the key constructs in this study.

Atmospheric touch point

Adopted from Jagadish (2013) and Mishra et al. (2014), A four-item scale is built to measure atmospheric touch point. The sample questions were “Their shopping environment increase my duration” and “Their shopping atmosphere impresses me” The items were measured on a 5-point Likert scale.

Communicative touch point

Four scales used in Sahin (2011) and Jagadish (2013) were modified and used in this study to measure communicative touch point. The sample questions are “I feel positive toward their advertising and promotions” and “Their advertising and promotions are informative”. The items were measured on a 5-point Likert scale
Product touch point

Product touch point is measured by four items adopted from Sahin (2011) and Jagadish’s (2013) study. The sample questions are “Their products are of high quality” and “Compared to other retailers/brands, they offer attractive prices”. The items were measured on a 5-point Likert scale.

Technological touch point

Four items from Stein and Ramaseshan (2015) and Jagadish (2013) were used to measure technological touch point. The sample questions are “Their technology makes shopping convenient” and “They have better technologies compared to other retailers/brands”. The items were measured on a 5-point Likert scale.

Process touch point

The Process touch point was measured by four items adopted from research by Stein and Ramaseshan (2015). The sample questions are “I can easily find information or products needed” or “Their service processing (shipping/returning) is convenient”. The items were measured on a 5-point Likert scale.

Employee-customer touch point

The Employee-customer touch point is measured based on the study from Rabbane et al. (2015) and Sahin (2011). The sample questions are “Their employees understand my needs”
and “Overall, their employees deliver excellent service”. The items were measured on a 5-point Likert scale.

**Customer-customer touch point**

Jagadish’s (2013) study was used to measure customer-customer touch point. Four items were used such as “Other customers’ opinion affected my opinion toward this retailer/brand” and “Overall, other consumer’s review about this retailer/brand is positive”. The items were measured on a 5-point Likert scale.

**Omni-channel customer experience**

According to Oh’s (2008) research, the general customer experience is measured from four items. The sample questions are “I am happy with the product I purchased from this retailer/brand” and “Overall, my experience with this retailer/brand is excellent”. The items were measured on a 5-point Likert scale.

### 3.3 Data analysis

#### 3.3.1 Research Objectives 1-8 data analysis methods

**Data screening and cleaning process**

Totally, the survey generated 348 responses (Appendix C: Survey data snippet), among which 19 responses were considered invalid due to incomplete responses on the part of the Survey participant. Further, we replace a missing value with the median value of all
responses of corresponding questions so that the other responses could be included in the data analysis.

**Exploratory Factor Analysis (EFA)**

Exploratory Factor Analysis (EFA) was conducted to investigate the dimensionality of the omni-channel customer experience. The recommended reliability scores alpha should be greater than 0.70 (Oh, 2008).

**Normality test and Confirmatory Factor Analysis (CFA)**

The basic assumption of CFA is the normality of the variables. Skewness and Kurtosis were used to test normality. The skewness value of zero indicates a normal distribution, while if the value is greater than three, it is considered an extremely skewed distribution.

Kurtosis value measures the extent of sharpness of the data. A Kurtosis value that is below 10 indicates a normal distribution in terms of its peakedness (Kline, 1998).

Confirmatory Factor Analysis (CFA) was then conducted to examine the adequacy of the hypothesized measurement relationship of the proposed model. Based on Kelloway’s recommendations (1998), three types of fitness indices – absolute, comparative and parsimonious fit index were used in the study to assess the measurement model fitness. The absolute fit was measured by the root mean square error of approximation (RMSEA) and $\chi^2$ statistics; the comparative fit was measured by comparative fit index (CFI); and parsimonious fit was measured by the results of parsimonious normed fit index (PNFI). If the
results meet the following criteria, then it indicates good fit of the proposed model in the study (Hu and Bentler, 1999):

- \((\chi^2)/\text{df} \leq 3\); adjusted goodness of fit (AGFI) \(\geq 0.80\);
- root mean square residual (RMR) \(\leq 0.1\);
- root mean square error of approximation (RMSEA) \(\leq 0.08\);
- comparative fit index (CFI) \(\geq 0.90\);
- goodness of fit (GFI) \(\geq 0.90\); and
- parsimonious normed fit index (PNFI) \(\geq 0.60\);

In order to make sure the items are free from error and consistent, the reliability is tested. A Cronbach’s alpha that is greater than 0.70 is an acceptable level of reliability.

This study tests three kinds of validity to make sure the items really reflect the latent variables: face validity, convergent validity, and discriminant validity. Face validity is established because all the items are adopted from previous literature. An average variance extracted (AVE) value that is greater or equal than 0.5 indicates convergent validity. Discriminated validity was tested by comparing the AVE value with squared correlations between any pair of constructs. A greater value of AVE than the correlations value indicates discriminate validity (Oh, 2008).
SEM

Last, Structural Equation Modeling (SEM) is conducted in SPSS AMOS to test the causal relationship among variables. Then objectives are discussed here based on SEM results.

Multi-group Analysis

To test the variations of different channels in the touch point based experience (TPE) model, multi-group analyses are conducted by SPSS AMOS. Measurement invariance test was conducted before multi-group mediation analysis.

3.3.2 Research Objective 9 analysis methods

The Wharton data set was used to test and evaluate Research Objective 9 and to investigate the role of RFM on predicting omni-channel consumers’ purchase probability for targeted promotions. Research Objective 9 is divided into two subset research objectives. The first research objective 9a is to examine the performance of this Fortune 500 Specialty Retailer’s targeted promotions. The second research objective 9b is to propose and evaluate a new method to predict consumer’s purchase probability so as to improve the performance of targeted promotions.

Objective 9a data screening and cleaning process

The Wharton data files are stored in a “raw” data format that is very similar to the way the fortune 500 specialty retailer sponsoring organization stores them. In particular, two data types were used in this project: “marketing data” and “purchasing data”.

61
The “marketing data” contains information of all direct marketing emails sent to the consumers from March 5, 2012 to March 2, 2014. Each row represents a household’s single exposure to email marketing, which include unique Household ID, Date, Email ID, as well as Email descriptions. There are 5% of rows that are exact duplicates because a household may have more than one email associated with the account. Totally over 3 million observations by 10 matrixes were read from “marketing data” file.

The “purchasing data” records all the transaction information from different channels. Each row represents a single line item purchased during a single transaction. One entry includes data for the Household ID, purchase details (such as price, quantity, discount, purchase date, purchase time), channels (online vs in-store), item details (item ID, description, sub-category, category, department), and store location (such as city, zip code, and region). There are over 11 million transactions with up to 40 data elements per transaction in this data file. The “marketing data” file and “purchasing data” file can be joint by the key variable “Household ID”.

Python and SAS were used during data cleaning and data analysis process considering the large size of the datasets. The raw datasets are in the format of comma separated values (CSV), where the frequency of errors is the first main challenge of this research. There are some typical errors in the dataset such as commas sporadically in the text entries or a single entry takes up several columns causing all data following it into incorrect columns. In an aggregate level, these errors are account for 5% of the rows in “marketing data” and about 10% of the rows in “purchasing data”.

62
We thus assume that these errors are not randomly distributed. In this case, the data is then affected in a non-random way, so an invalid result will be reported if these error rows are deleted.

The data was then cleaned through Python. The “purchasing data” is more difficult to clean because of larger scales, more frequent errors, and more varied location of errors. The entry errors and missing values in “purchasing data” make it not feasible to build model based on dollar value variables or location variables. However, two key variables were not affected by these errors, which are the date they made a purchase and the household IDs. Combining cleaned “marketing data” with cleaned “purchasing data” through key variable “Household ID”, we can know: whether the customer made a purchase on each date, the quantity and types of emails the customer received on this day and preceding days.

**Research Objective 9a data analysis model**

To examine the performance of the Fortune 500 Specialty Retailer’s targeted promotions, logistic regression was first used to model the probability of a customer making a purchase on a certain date after their recent receipt of emails. The logistic regression model is show as follows in Equation 4:
\[
\hat{Y}_{ij} = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ij})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ij})} \quad \text{for } i = 1, \ldots, 60000 \quad \text{and } j = 1, \ldots, 722
\]

where

\[
\hat{Y}_{ij} \text{ represents the probability of customer } i \text{ making a purchase on day } j
\]

\[
x_{1i} = 1 \text{ if customer } i \text{ has a registered email address with the company}
\]

\[
x_{2ij} = 1 \text{ if customer } i \text{ received an email on day } j
\]

\[
x_{3ij} = \text{the number of emails customer } i \text{ received on days } j - 7 \text{ to } j - 1
\]

The intercept term \( \beta_0 \) is included in the model because customers have some baseline probability to purchase products even though they have never received an email.

Considering that the population of customers who give their email addresses to the company might be more likely to make purchases, we include parameter \( \beta_1 \) in the model to indicate whether the customer had a registered email. \( \beta_2 \) and \( \beta_3 \) are parameters for the change in purchase probability that comes from receiving an email on the purchase date and on the days leading up a week to it, respectively.
However, simple logistic regression is not appropriate in the research since the data is longitudinal and correlated by customer and over time, which violate one key assumption in regular binary logistic regression that observations are independent of each other.

There are two types of statistical models being widely used to model binary data with correlation. The first one is the marginal or population-averaged model, where the correlation of the measurements is accounted through a robust covariance matrix and a group effect is not explicitly included in the model. One popular method is a non-likelihood or quasi-likelihood based method - Generalized Estimating Equations (GEE), which produces consistent estimators of the regression parameters and their variances.

Another type of model to deal with correlated binary data is the conditional or subject-specific model, where the dependence of observations were modeled by including a group specific effect in the model. Generalized linear mixed model (GLMM) is one of these conditional models that include random effects, which treats the group effects as random and uses maximum likelihood methods for estimation. The difference between marginal and conditional model is that marginal models such as GEE measures the effects average across all subjects, while conditional models such as GLMM measures the effect for a particular subject (Despa, n.d.).

In this study, Research Objective 9a aims to identify the average improve rate of purchase probability under the influence of targeted promotions, while Research Objective 9b focuses on understanding the role of consumers’ behavior characteristics such as RFM in predicting
consumers' purchase probability. Thus, the GEE model was employed in research objective 9a and GLMM model was used in research objective 9b.

Using “proc genmod” in SAS, this study specifies "working independence" by customer and within date. Standard Errors were estimated by specifying model-based standard errors, which allows the specification of an assumed correlation structure to account for the non-independence of observations, as opposed to empirical-based standard errors that assumes independence of observations.

**Research Objective 9b data screening and cleaning process**

To narrow down the research scope to omni-channel consumers, we set the filtering criteria as known household customers (remove professional buyers) whose online shopping frequencies are larger or equal to 10 times. The filtered dataset consists of 156,768 transactions over 2 years, year 2012 and 2013. A total of 274 households (customers) are included and all products are classified to more than 205 categories. Each transaction includes an identifier id for the household, the date, the quantity of product, the price, and specific code for the item’s category.

During our data cleaning and validation procedures, some corrupted and misplaced data entries were detected and corrected. In addition, some of the transactions have a negative value in transaction quantity because they represent customers’ returning products. This study simply deleted the negative quantity transactions from the dataset, which accounts for only 4%, because of the difficulties to match the returns transaction with the original
transaction and it is unnecessary to delete the corresponding original transactions since they represent consumers original purchase demand.

**Research Objective 9b data analysis model**

The objective is to predict the possibility of a certain customer (household) purchase a product in a certain category (total of 205 categories) or class in a certain period of time. The study establishes a predictive model based on the RFM model. As discussed earlier in Chapter 2, R represents the number of days between the last transaction and the current transaction for this customer, specifically in this study, the number of days between the last transaction and first day of the current month for a customer. F is the average number of transaction that happened per month starting from the first transaction month in the dataset. M is the total amount of money spent on this kind of grocery until the current month.

Thus for consumer or household i, class j and time t, 1 ≤ i ≤ N, 1 ≤ j ≤ C, this study built a binary variable $Y_{ijt}$ such that $P(Y_{ijt} = 1)$ is a function of $R_{ijt}$, $F_{ijt}$, $M_{ijt}$. Then $Y_{ijt}$ was set to 1 if there is any purchase made in current month and to 0 if no purchase was made. N is the total number of customers and C is the total number of product classes. Considering the definition of recency, there is not enough data to calculate recency, thus this study uses the data that range from month 7 to 18 as training data and month 19 to 24 as test dataset to evaluate the model. The following variables are required to calculate RFM:

- Household id: unique identifier for the household
- Transaction date: the date when a transaction happens
• Unit quantity: the quantity of product

• Unit sales: price per unit

• Item class code: code for the items class

• Item class description: description of item class

• Transaction month: the month when a transaction happens

• Transaction month number: number of month from March 2012

From the variables above, we would calculate the following variables for prediction model:

• \( Y_{ijt} \): Indicator variable for purchase of customer i on products from class j.

• \( R_{ijt} \): The number of days between the last transaction and first day of the current month for customer i and products from class j.

• \( F_{ijt} \): The average number of transaction happened per month from the beginning of the year 2012 to current month for customer i and class j.

• \( M_{ijt} \): The total amount of money spent on this kind of products from the beginning of the year 2012 to current month for customer i and class j.

After all RFM actual values are calculated, all values were normalized within categories of products. The normalized version was based on equation 5 below. Since the minimum value
of RFM is always zero in the dataset, the normalized equation becomes the original value divided by the maximum of that variable within the class.

\[
RFM_n = \frac{RFM - RFM_{\text{min}}}{RFM_{\text{max}} - RFM_{\text{min}}}
\]  
(5)

where:

\(RFM_n\) is normalized recency, frequency, monetary.

\(RFM\) is original recency, frequency, monetary.

\(RFM_{\text{max}}\) is the maximum value of recency, frequency, monetary.

\(RFM_{\text{min}}\) is the minimum value of recency, frequency, monetary.

As mentioned in the previous chapter, AHP method that was first proposed was not a reliable method to deal with large dimensional data set. Since there are over 200 classes in the data set, and each classes will be used to build targeted email campaigns, which requires over \(n*(n-1)/2 = 19900\) times pairwise comparison (Teknomo, 2012). It is impossible to get reliable ranking results due to the variation in the comparison. Also the core value of AHP is pairwise comparison of multi-criteria items that cannot easily be compared. In this study normalization of different criteria was conducted to measure and compare the multi-criteria. Thus, AHP method was thus abandoned in this study.
This study used a generalized linear mixed effect models for each product class (total of 205 classes) to fit the data. For household i, class j and time t, $1 \leq i \leq N$, $1 \leq j \leq C$, the jth model was set to be as shown in Equation 6.

$$\text{logit}\{P(Y_{ijt}=1|R_{ijt}, F_{ijt}, M_{ijt})\} = \beta_{0,j} + \beta_{R,j} R_{ijt} + \beta_{F,j} F_{ijt} + \beta_{M,j} M_{ijt}$$

(6)

where

$Y_{ijt}$: Indicator variable for purchase of customer i on products from class j.

$R_{ijt}$: The number of days between the last transaction and first day of the current month for customer i and products from class j.

$F_{ijt}$: The average number of transaction that occurred per month from the beginning of the year 2012 to current month for customer i and class j.

$M_{ijt}$: The total amount of money spent on this kind of products from the beginning of the year 2012 to current month for customer i and class j.

The logit function is $\text{logit}(x) = \log(x/x-1)$. The quantity modeled is the probability of household i making at least one purchase in product class j during monthly time period t given the corresponding recency, frequency and money information. The coefficients $\beta_{0,j}, \beta_{R,j}, \beta_{F,j}, \beta_{M,j}$ on the right side of the equation are fixed effects of RFM.
The model fitting results indicate that most estimated coefficients for M have very high variance. One possible explanation is that for a certain class, the household with higher income, more money spent indicates higher demand and purchase probability, whereas for lower income household, they might consider spending less money on things that they already bought. Bagheri and Tarokh (2014) clustered the customer into two groups to take into account the difference of consumption behavior among customers. Hence, to deal with the high variance on M and make the model more flexible, a generalized mixed effect model was established with a random effect for households on Money as show in equation 7, where $\beta_{M,j}$ is fixed and $B_{M,i,j}$ is random by household i and has zero mean.

$$\text{logit}\{P(Y_{ijt}=1|R_{ijt},F_{ijt},M_{ijt})\} = \beta_{0,j} + \beta_{R,j}R_{ijt} + \beta_{F,j}F_{ijt} + (\beta_{M,j} + B_{M,i,j})M_{ijt}$$  \hspace{1cm} (7)$$

$$B_{M,i,j} \sim i.N(0,\sigma^2_M)$$

where

$Y_{ijt}$: Indicator variable for purchase of customer i on products from class j.

$R_{ijt}$: The number of days between the last transaction and first day of the current month for customer i and products from class j.

$F_{ijt}$: The average number of transaction happened per month from the beginning of the year 2012 to current month for customer i and class j.
\(M_{ij}\): The total amount of money spent on this kind of products from the beginning of the year 2012 to current month for customer i and class j.

Based on the model estimated by the training data, it is possible to predict the probability of purchasing per customer, category and month for the testing data set. Specifically, the “glmer” function from the “lme4” package in R Studio was used. Then, the model was evaluated by plotting an ROC (Receiver Operation Characteristic) curve, which shows how false positive rate (FPR) and true positive rate (TPR) is affected by different cutoff value. The area under the curve (AUC) is often used to evaluate the performance of prediction for binary outcomes, a large AUC value (≥0.8) indicates an excellent predict performance (Lantz, 2013).

Comparison of the prediction result of the proposed model with the prediction result of a baseline model was performed, which is a traditional logistic regression model assuming that the fixed effect of R, M and F are the same across all households and categories of products. The baseline model is illustrated as in the equation 8:
\[
\text{logit}\{P(Y_{ijt} = 1|R_{ijt}, F_{ijt}, M_{ijt})\} = \beta_0 + \beta_R R_{ijt} + \beta_F F_{ijt} + \beta_M M_{ijt}
\]

(8)

where

\(Y_{ijt}\): Indicator variable for purchase of customer \(i\) on products from class \(j\).

\(R_{ijt}\): The number of days between the last transaction and first day of the current month for customer \(i\) and products from class \(j\).

\(F_{ijt}\): The average number of transactions occurring per month from the beginning of the year 2012 to current month for customer \(i\) and class \(j\).

\(M_{ijt}\): The total amount of money spent on this kind of products from the beginning of the year 2012 to current month for customer \(i\) and class.
A. Research Objectives 1-8 survey data results

4.1 Descriptive analysis

After deleting those incomplete responses where less than 20 percent were finished, we have 329 valid response data. Among the 329 respondents, 58.7% were female and the age range was from 18 to 26 for the majority (82.1%) of the respondents. The survey asked the respondents to select the channel they were using to search, compare, or buy apparel or footwear products. This channel choice result shows that 31.3% of the respondents shop apparel or footwear products using only one channel, in either store or online. While 68.7% of the respondents leverage two or more channels to shop apparel or footwear products, 39.2% use three or more channels, such as in store, online, and on mobile devices (Table 1).
Table 1: Descriptive statistic of gender and shopping channel

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>134</td>
<td>40.7</td>
</tr>
<tr>
<td>Female</td>
<td>193</td>
<td>58.7</td>
</tr>
<tr>
<td>Unknown</td>
<td>2</td>
<td>.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shopping Channel</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>82</td>
<td>24.9</td>
</tr>
<tr>
<td>B</td>
<td>19</td>
<td>5.8</td>
</tr>
<tr>
<td>A^B</td>
<td>63</td>
<td>19.1</td>
</tr>
<tr>
<td>A^C</td>
<td>6</td>
<td>1.8</td>
</tr>
<tr>
<td>B^C</td>
<td>28</td>
<td>8.5</td>
</tr>
<tr>
<td>A^B^C</td>
<td>127</td>
<td>38.6</td>
</tr>
<tr>
<td>A^B^C^D</td>
<td>2</td>
<td>.6</td>
</tr>
</tbody>
</table>

* A means in store; B means on computer; C means on mobile device; D means others (TV or catalogs); A^B means both in store and on computer, etc;
4.2 Exploratory Factor Analysis (EFA)

Factor Analysis is conducted as a multivariate statistical approach to identify a set of latent constructs that are capable to explain a large portion of the total variances in the items (Chandrapal & Brahmbhatt, 2015). EFA is a dimension reduction method to discover the factor structure and examine its internal reliability, which are often used when the dimensions of a measure is not clear. The goal of the EFA is to reduce and identify the number of constructs, where correlated constructs are grouped together to form a new aggregated variable, and detect the underlying factor structures (Suhr, 2006).

**Sampling adequacy test**

Before the exploratory factor analysis, a Kaiser-Meyer-Olkin Measure of sampling adequacy was conducted to ensure the suitability of factor analysis. The recommended KMO value for a “good adequacy” is between 0.7 and 0.8, and a “great adequacy” with KMO greater than 0.8 (Kaiser, 1974). Bartlett’s Test of Sphericity is another indicator to test the significance of the study, which reflects the strength of the relationship among variables (Chandrapal & Brahmbhatt, 2015). The results show a Kaiser-Meyer-Olkin Measure of Sampling Adequacy value of 0.828 and a significant level of Bartlett's Test of Sphericity, which confirmed the adequacy of the following EFA procedure.
Factor analysis

Then, exploratory factor analysis (EFA) was conducted with maximum likelihood extraction method and Varimax with Kaiser Normalization rotation method. EFA results (Table 2) suggest eight factors with initial eigenvalue greater than 1 even though no particular number of factors was assigned before factor analysis. Two items are removed due to their low factor loading (<0.3). Together, these eight factors explained 60% of the total variance.
Table 2: Summary of Exploratory Factor Analysis Results

<table>
<thead>
<tr>
<th>Items \ Factors</th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
<th>Factor4</th>
<th>Factor5</th>
<th>Factor6</th>
<th>Factor7</th>
<th>Factor8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etoc2</td>
<td>0.800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etoc4</td>
<td>0.759</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etoc1</td>
<td>0.748</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etoc3</td>
<td>0.744</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech2</td>
<td></td>
<td>0.842</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech1</td>
<td></td>
<td>0.832</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech3</td>
<td></td>
<td>0.705</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech4</td>
<td></td>
<td>0.549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall1</td>
<td></td>
<td></td>
<td>0.734</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall4</td>
<td></td>
<td></td>
<td></td>
<td>0.727</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.721</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commu2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.762</td>
<td></td>
</tr>
<tr>
<td>Commu4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.729</td>
</tr>
<tr>
<td>Commu1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.703</td>
</tr>
<tr>
<td>Commu3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.674</td>
</tr>
<tr>
<td>Atmo3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.792</td>
</tr>
<tr>
<td>Atmo4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.757</td>
</tr>
<tr>
<td>Atmo2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.616</td>
</tr>
<tr>
<td>Atmo1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.601</td>
</tr>
<tr>
<td>Product3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.698</td>
</tr>
<tr>
<td>Product2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.696</td>
</tr>
<tr>
<td>Product4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.648</td>
</tr>
<tr>
<td>Product1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.623</td>
</tr>
<tr>
<td>Ctoc1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.767</td>
</tr>
<tr>
<td>Ctoc3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.709</td>
</tr>
<tr>
<td>Ctoc2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.647</td>
</tr>
<tr>
<td>Ctoc4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.276</td>
</tr>
<tr>
<td>Process4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.881</td>
</tr>
<tr>
<td>Process3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.830</td>
</tr>
<tr>
<td>Process2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.570</td>
</tr>
<tr>
<td>Process1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.284</td>
</tr>
</tbody>
</table>

CRONBACH’s a: 0.884 0.884 0.866 0.830 0.840 0.769 0.774 0.751  
Eigenvalues: 7.352 3.031 2.620 2.298 1.650 1.482 1.327 1.275

Extraction method: Maximum Likelihood;
Rotation Method: Varimax with Kaiser Normalization.
4.3 Confirmatory Factor Analysis (CFA)

To ensure reliability and validity of the constructs and to examine the adequacy of the hypothesized measurement relationship of the proposed model, CFA was first conducted, followed by a structural equation model to test the relationships between constructs, as suggested by Anderson and Gerbing (1988).

**Normality Assumption test**

Skewness and Kurtosis were examined to test whether the data meet the multivariate normality assumption of CFA. Skewness measures the asymmetry of a distribution where “Positive Skew” values indicate a right skewed distribution and “Negative Skew” values indicate a left skewed distribution. A normal distribution will generate a skew value of zero. Kurtosis reflects the peakedness of a distribution. Positive kurtosis indicates high peak and negative kurtosis means flat-topped top distribution (Kim, 2013). Skewness and kurtosis values between -2 and +2 are suggested as acceptable values to prove normal distribution (George & Mallory, 2010). The kurtosis and skewness results show that all values ranging between -1.34 to +1.81, which ensures the normality assumptions of CFA.

**CFA Model fit**

Then CFA was conducted to test the model fit to the survey data. According to Kelloway’s recommendations (1998), three types of fitness indices are used as the measurement model fitness index: absolute, comparative and parsimonious fit index. The absolute fit was indicated by the root mean square error of approximation (RMSEA) and $\chi^2$ statistics; the
comparative fit is measured by comparative fit index (CFI); and parsimonious fit is measured by the results of parsimonious normed fit index (PNFI). Based on Hu and Bentler’s recommendation (1999), \( (\chi^2)/df \leq 3 \), adjusted goodness of fit (AGFI) ≥ 0.80, root mean square residual (RMR) ≤ 0.1, root mean square error of approximation (RMSEA) ≤ 0.08, and comparative fit index (CFI) ≥ 0.90 and goodness of fit (GFI) ≥ 0.90, and parsimonious normed fit index (PNFI) ≥ 0.60 are used as indications of acceptable model fit in this study.

The CFA measurement model fit results shown in Table 3 indicate that all model fit indices meet the recommendation criteria, which indicates a good model fit to the data.
Table 3: CFA model fit indices

<table>
<thead>
<tr>
<th>Type of Fit Index</th>
<th>Fit Indices</th>
<th>Recommendation</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>CMIN/DF (Minimum discrepancy/degree of freedom)</td>
<td>≤ 3</td>
<td>1.396</td>
</tr>
<tr>
<td></td>
<td>GFI (Goodness of fit)</td>
<td>≥ 0.90</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>AGFI (Adjusted goodness of fit)</td>
<td>≥ 0.80</td>
<td>0.880</td>
</tr>
<tr>
<td></td>
<td>RMSEA (Root mean square error approximation)</td>
<td>≤ 0.08</td>
<td>0.035</td>
</tr>
<tr>
<td>Comparative</td>
<td>CFI (Comparative fit index)</td>
<td>≥ 0.90</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>NFI (Normed fit index)</td>
<td>≥ 0.90</td>
<td>0.910</td>
</tr>
<tr>
<td>Parsimonious</td>
<td>PNFI (Parsimonious normed fit index)</td>
<td>≥ 0.60</td>
<td>0.691</td>
</tr>
</tbody>
</table>

**Reliability**

Coefficient alpha was calculated to assess the reliability of each construct. Coefficient alpha score exceeding 0.70 is considered as an acceptable level of reliability (Nunnally & Bernstein, 1994). According to Table 2, Cronbach’s alpha for all eight constructs are above 0.7, which indicated a good internal consistency reliability. Also, composite reliability (CR) is
calculated to further examine the overall reliability of the measurement in the model. The results show that all CR values are greater than the recommended minimum level of 0.7 (Nunnally & Bernstein, 1994).

**Validity**

Convergence validity of the latent constructs is assessed by the average variance extracted (AVE). As shown in Table 4, all the constructs had AVE values higher than 0.50 as recommended (Fornell and Larcker, 1981) except for the process touch point (0.493) and product touch point (0.450) constructs. However, Malhotra and Dash (2011) pointed out that AVE is a more conservative and stricter measure than CR, and the researcher may conclude the adequate convergent validity of the construct based on CR alone. As discussed before, the CR value of process and product touch point meet the criteria. Together, the AVE and CR demonstrate adequate convergence validity.

Discriminant validity is firstly measured by comparing the squared root of AVE with correlations between the constructs. As recommended (Table 5), squared root of AVE is greater than constructs correlations for all constructs. (Fornell and larcker, 1981). Also, the Maximum Shared Variance (MSV) is compared with Average Variance Extracted (AVE). The result (Table 4) shows that MSV is smaller than AVE for all constructs, which further confirms discriminant validity (Hair. etc., 2010).
Table 4: CFA validity results

<table>
<thead>
<tr>
<th></th>
<th>CR (composite reliability)</th>
<th>AVE (Average Variance Extracted)</th>
<th>MSV (Maximum Shared Variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>process</td>
<td>0.713</td>
<td>0.493</td>
<td>0.281</td>
</tr>
<tr>
<td>Product</td>
<td>0.710</td>
<td>0.450</td>
<td>0.107</td>
</tr>
<tr>
<td>commu</td>
<td>0.831</td>
<td>0.552</td>
<td>0.131</td>
</tr>
<tr>
<td>atmo</td>
<td>0.833</td>
<td>0.559</td>
<td>0.298</td>
</tr>
<tr>
<td>ctoc</td>
<td>0.777</td>
<td>0.538</td>
<td>0.173</td>
</tr>
<tr>
<td>tech</td>
<td>0.885</td>
<td>0.658</td>
<td>0.281</td>
</tr>
<tr>
<td>etoc</td>
<td>0.885</td>
<td>0.659</td>
<td>0.203</td>
</tr>
<tr>
<td>overall</td>
<td>0.862</td>
<td>0.610</td>
<td>0.298</td>
</tr>
</tbody>
</table>
Table 5: Factor correlation matrix with square root of the AVE on bold number

<table>
<thead>
<tr>
<th></th>
<th>process</th>
<th>product</th>
<th>commu</th>
<th>atmo</th>
<th>ctoc</th>
<th>tech</th>
<th>etoc</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>process</td>
<td>0.627</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>product</td>
<td>0.195</td>
<td>0.671</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>commu</td>
<td>0.258</td>
<td>0.369</td>
<td>0.743</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>atmo</td>
<td>0.414</td>
<td>0.416</td>
<td>0.546</td>
<td>0.748</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ctoc</td>
<td>0.451</td>
<td>0.318</td>
<td>0.438</td>
<td>0.083</td>
<td>0.733</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tech</td>
<td>0.222</td>
<td>0.082</td>
<td>0.257</td>
<td>0.291</td>
<td>-0.188</td>
<td>0.811</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etoc</td>
<td>0.264</td>
<td>0.530</td>
<td>0.353</td>
<td>0.127</td>
<td>-0.107</td>
<td>0.430</td>
<td>0.812</td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>0.257</td>
<td>0.340</td>
<td>0.361</td>
<td>-0.292</td>
<td>0.394</td>
<td>-0.314</td>
<td>0.549</td>
<td>0.781</td>
</tr>
</tbody>
</table>
4.4 Structural Equation Modeling

A structural equation modeling is conducted in SPSS and AMOS (v24) to test the direct effect of the proposed seven touch points (atmospheric, technological, communicative, process, employee-customer, customer-customer, and product) on overall consumer experience. The SEM model fit indices are at acceptable level as shown in Table 6 (CMIN/DF=1.586, GFI=0.897, AGFI=0.865, RMSEA=0.042, CFI=0.953, NFI=0.901, PNFI=0.718).

Table 6: SEM model fit

<table>
<thead>
<tr>
<th>Type of Fit Index</th>
<th>Fit Indices</th>
<th>Recommendation</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>CMIN/DF (Minimum discrepancy/degree of freedom)</td>
<td>≤ 3</td>
<td>1.586</td>
</tr>
<tr>
<td></td>
<td>GFI (Goodness of fit)</td>
<td>≥ 0.90</td>
<td>0.919</td>
</tr>
<tr>
<td></td>
<td>AGFI (Adjusted goodness of fit)</td>
<td>≥ 0.80</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>RMSEA (Root mean square error approximation)</td>
<td>≤ 0.08</td>
<td>0.042</td>
</tr>
<tr>
<td>Comparative</td>
<td>CFI (Comparative fit index)</td>
<td>≥ 0.90</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>NFI (Normed fit index)</td>
<td>≥ 0.90</td>
<td>0.901</td>
</tr>
<tr>
<td>Parsimonious</td>
<td>PNFI (Parsimonious normed fit index)</td>
<td>≥ 0.60</td>
<td>0.718</td>
</tr>
</tbody>
</table>
SEM path analysis results (Table 7) suggest that five touch points (Atmospheric, Communicative, Process, Employee-Customer, and Product) have significant effect on overall customer experience with a R square equal to 0.47. However, Technological and Customer-Customer touch point indicates non-significant p values when measuring their direct effects on overall customer experience. Thus, we could not make a conclusion of the significant influence of these two constructs on overall customer experience.
<table>
<thead>
<tr>
<th>Experience</th>
<th>Standardized coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmospheric</td>
<td>.371</td>
<td>***</td>
</tr>
<tr>
<td>Technological</td>
<td>-.085</td>
<td>.273</td>
</tr>
<tr>
<td>Communicative</td>
<td>.239</td>
<td>***</td>
</tr>
<tr>
<td>Process</td>
<td>.260</td>
<td>***</td>
</tr>
<tr>
<td>Employee-customer</td>
<td>.149</td>
<td>.029</td>
</tr>
<tr>
<td>Customer-customer</td>
<td>-.015</td>
<td>.815</td>
</tr>
<tr>
<td>Product</td>
<td>.191</td>
<td>***</td>
</tr>
</tbody>
</table>
4.5 Multi-group Analysis

Research Objective 8 is to investigate the influence of consumers’ channel choice (single channel VS multi-channel) on the relationship between touch points and omni-channel consumer experience. In this case, channel choice plays a moderation role between the touch points and overall customer experience. Multi-group analysis was conducted to measure the moderation effect. Before multi-group analysis, test of invariance was conducted to make sure the proposed model and structure is statistically invariant between multi-groups, in this study, two groups: single channel and multi-channel. In other word, we have to make sure the model structure is same for two groups, just like same scale for two rulers to measure two things.

Test of Invariance

To assure that the comparisons of the latent variable and their relationships are valid across groups, test of measurement invariance and test of structural invariance are assessed through SPSS and AMOS Multi-group analysis function.

First of all, configural invariance is tested by running a multi-group CFA to ensure the basic model structure is conceptualized in the same way across groups. The model fit results (CMIN/DF=1.53, GFI=0.893, AGFI=0.887, RMSEA=0.043, CFI=0.947, NFI=0.914, PNFI=0.684) suggest a good model fit and confirm the established configural invariance, which ensure the following invariance to be meaningful. Then the model is tested by gradually increasing constraints (Unconstrained model → Equal Measurement Weight →
Equal Structural Weight $\rightarrow$ Equal Structural residual and covariance $\rightarrow$ Equal Measurement residuals). Comparing the $\chi^2$ and df with the previous less restricted model will yield a set of $\Delta \chi^2$ and $\Delta$df. When the equal factor loading across single channel and multi-channel group was constrained, the model generated a significant $p$ value, which indicates non-invariance across groups. However, full measurement invariance is too strict and unrealistic in practice (Milfont & Fischer, 2010). Vandenberg & Lance (2000) suggest that in order to continue the sequent tests, at least partial metric invariance must be achieved. Partial measurement invariance requests only parts of the parameters are constrained while other parameters are allowed to vary in the model across groups. After releasing the equal constrains of three factors loadings that generate the highest difference, we get statistically non-significant $p$ values (Table 8) across two groups with all levels of constraints enforced to the model, which gives up adequate permission to perform the multi-group moderation testing.
Table 8: Test of invariance results

<table>
<thead>
<tr>
<th>Overall Model</th>
<th>Chi-square</th>
<th>df</th>
<th>p</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained</td>
<td>992.818</td>
<td>602</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal Measurement Weight</td>
<td>1021.660</td>
<td>622</td>
<td>0.09</td>
<td>Partial Invariant</td>
</tr>
<tr>
<td>(Equal Factor loading)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal Structural Weight</td>
<td>1030.103</td>
<td>629</td>
<td>0.295</td>
<td>Invariant</td>
</tr>
<tr>
<td>Equal Structural residual covariance</td>
<td>1062.833</td>
<td>652</td>
<td>0.086</td>
<td>Invariant</td>
</tr>
<tr>
<td>Equal Measurement residuals</td>
<td>1126.852</td>
<td>707</td>
<td>0.189</td>
<td>Invariant</td>
</tr>
</tbody>
</table>
**Multi-group moderation analysis**

The multi-group moderation analysis results (Table 9) show that for single channel (In store or Online) group, four touch points (atmospheric, process, employee-customer, product) are found to have significant influence on overall customer experience. For multi-channel group, four touch points (atmospheric, communicative, process, product) are also found to have significant influence on overall customer experience. Comparing the two groups (single channel group and multi-channel group), we could find that the path between communicative and overall customer experience is non-significant for single channel group and significant for multi-channel group. Also, the path between employee-customer and overall customer experience is significant for single channel group and become non-significant in multi-channel group. Technological and customer-customer touch points remain non-significant influence on overall customer experience for both groups.
Table 9: Multi-group moderation analysis results

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Single Channel Path</th>
<th>Multi-Channel Path</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standardized coefficient</td>
<td>P</td>
</tr>
<tr>
<td>Experience</td>
<td>Atmospheric</td>
<td>.571</td>
<td>** .002**</td>
</tr>
<tr>
<td>Experience</td>
<td>Technological</td>
<td>-.099</td>
<td>.269</td>
</tr>
<tr>
<td>Experience</td>
<td>Communicative</td>
<td>.119</td>
<td>.386</td>
</tr>
<tr>
<td>Experience</td>
<td>Process</td>
<td>.189</td>
<td>** .046**</td>
</tr>
<tr>
<td>Experience</td>
<td>Employee-customer</td>
<td>.180</td>
<td>** .045**</td>
</tr>
<tr>
<td>Experience</td>
<td>Customer-customer</td>
<td>.005</td>
<td>.671</td>
</tr>
<tr>
<td>Experience</td>
<td>Product</td>
<td>.311</td>
<td>** .002**</td>
</tr>
</tbody>
</table>

** denotes p value less than 0.01
*** denotes p value less than 0.001

There are many reasons that could cause the non-significant effect for single channel and multichannel groups, which may include the millennial sample characteristics, consumer behavior changed by mobile and online technologies etc. Specific explanations will be provided in the Conclusion and Discussion Chapter. The next section provides results of the Wharton data analysis.
B. Research Objective 9 Wharton data results

4.6 Research Objective 9a results

Research Objective 9a is to examine the performance of this Fortune 500 Specialty Retailer’s targeted promotions. GEE model was used to measure the overall emails performance as well targeted emails performance.

GEE model converges to the parameter estimates that are showed given in Table 10 below. All of the coefficients of the logistic function are highly significant. The intercept term $\beta_0$ that indicates the baseline probability is -3.12, $\beta_1$ that indicates whether the customer registered or not is 0.66, and $\beta_2$ and $\beta_3$ that indicates the change in purchase probability that comes from receiving an email on the purchase date and on the days leading up a week to it, is 0.12 and 0.037 respectively.

However, the nature of the logistic model makes it difficult to interpret what the effects or coefficients actually represent. To better understand the results, we transform the logit probability to actual probability. Table 11 illustrates the actual and increasing in purchase probability that comes from changing the value of the variables, correspondently.
Table 10: Parameter estimates with all emails

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.1176</td>
<td>0.0009</td>
<td>-3.1193 -3.1159</td>
<td>-3577.4</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>reg</td>
<td>0.6585</td>
<td>0.0017</td>
<td>0.6551 0.6618</td>
<td>386.39</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>email</td>
<td>0.1242</td>
<td>0.0028</td>
<td>0.1187 0.1298</td>
<td>43.83</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>0.0377</td>
<td>0.0007</td>
<td>0.0364 0.0391</td>
<td>53.54</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

Note: Intercept: baseline probability without registering and receiving emails; Reg: register; Email: receive email on purchase day; Week: receive 1-6 emails one week before purchase day; All parameter are significant at 95% confident level.

Table 11: Purchase probability given the receipt values of all emails

<table>
<thead>
<tr>
<th>Registered Email Received on purchase date</th>
<th>Number received the Week Before</th>
<th>Probability of Purchase</th>
<th>Marginal Increase in Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>4.2%</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>7.9%</td>
<td>3.6%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>8.8%</td>
<td>0.9%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9.1%</td>
<td>0.3%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9.5%</td>
<td>0.3%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10.1%</td>
<td>0.3%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10.5%</td>
<td>0.3%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10.8%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>
From Table 11, we could find that the baseline purchase probability is 4.2%, and those who registered or subscribed the emails are 3.6% more likely to purchase compare to baseline consumers who do not registered. Further, the customers who received emails on purchase date will increase 0.9% purchase probability (become 8.8%). The more emails the consumer received the week before (start from the purchase date), the more likely they will make a purchase (0.3% increase per email). When they received 6 emails the week before, which means they received email everyday, their purchase probability become 10.8%. This is 2.9% more probability comparing to those customers who did not receive emails. The marginal increase in probability may appear small, this should be considered in the context of the scale of the data. An increase of 2.9% in purchase probability across nearly 60000 customers will result in 1740 additional purchase per day. The results reveals that this Fortune 500 Specialty Retailer’s direct email campaigns have a significant influence on consumer behavior.

To answer the question of whether targeted emails affect consumer purchase probability more than other marketing emails, this research conducted the likewise analysis using only the emails that make recommendations to the customer based on previous purchases. These targeted emails are in the form of “top 10 items recommend for you”. Thus, a new dataset was generated excluding non-targeted emails. The baseline purchase rate and registration effect are expected to be increased to accommodate the effects of non-targeted emails since non-targeted emails are not counted in the model. In other words, \( \beta_0 \) and \( \beta_1 \) will be different. If the effects of emails stay same, i.e. if \( \beta_2 \) and \( \beta_3 \) do not change, this would indicate that targeted emails have no more influence than the overall generic marketing emails. Another
factor to point out is that due to the rarity of the targeted emails (only 1700 out of 3.4 million), the targeted emails model only includes the effect of receiving one email on the purchase date or in the week before rather than 6. The parameter estimates and purchasing probability change for targeted email model are given in Tables 12 and 13, respectively.

Table 12: Parameter Estimates with targeted emails only

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z    | Pr > |Z| |
|-----------|---------|----------------|-----------------------|------|-------|---|
| Intercept | -2.8980 | 0.0007         | -2.8993 -2.8966       | -4200.8 | <.0001 |
| reg       | 0.6903  | 0.0036         | 0.6831 0.6974         | 189.43 | <.0001 |
| email     | 0.6808  | 0.0636         | 0.5561 0.8054         | 10.70 | <.0001 |
| week      | 0.5727  | 0.0267         | 0.5204 0.6250         | 21.45 | <.0001 |

Note*: Intercept: baseline probability without registering and receiving emails; Reg: register; Email: receive email on purchase day; Week: receive 1 email one week before purchase day; All parameter are significant at 95% confident level.
Table 13: Purchase probability given the receipt values of targeted emails

<table>
<thead>
<tr>
<th>Registered</th>
<th>Email Received on purchase date</th>
<th>Number received the Week Before</th>
<th>Probability of Purchase</th>
<th>Marginal Increase in Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.2%</td>
<td>5.2%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9.9%</td>
<td>4.7%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>17.8%</td>
<td>7.9%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>16.3%</td>
<td>6.4%</td>
</tr>
</tbody>
</table>

As we expected, Table 12 shows that all effects are significant. The intercept term $\beta_0$ that indicates the baseline probability is -2.898, $\beta_1$ that indicates whether the customer registered or not is 0.69, and $\beta_2$ and $\beta_3$ that indicates the change in purchase probability that comes from receiving an email on the purchase date and on the days leading up a week to it, is 0.68 and 0.57 respectively. Likewise, we then transform the logit probability to actual probability. The results in Table 13 indicates that the baseline purchase probability is 5.2%, and those who registered or subscribed the emails are 4.7% more likely to purchase compare to baseline consumers who do not registered. This results match with our assumption that the baseline probability and registered effects will be slightly higher because of the exclusion of non-targeted emails. Also, results show that the consumers who received one targeted emails on the purchase date or received one targeted emails the week before purchase date are 7.9% and 6.4% more likely to make a purchase. Thus we could make a conclusion that targeted emails dramatically increase consumer purchase probability than overall generic emails (0.9% and 0.3% lift) (Figure 7).
Figure 7: Comparison of overall generic emails and targeted emails effects
4.7 Research Objective 9b results

Research Objective 9b is to propose and evaluate a new method to predict consumer’s purchase probability so as to improve the performance of targeted promotions. GLMM (Generalized Linear Mixed Model) was used to model the influence of RFM on consumers’ purchase probability. Figure 8, 9, 10 show the distribution of estimated fixed effect for Recency, Frequency and Money, across product categories, thus x-axis in Figure 8, 9, 10 are categories numbers, for example, 1 means category number 1.

The histogram for fixed effects of Recency across classes shows a mode around -2, where a negative coefficient suggests that the shorter time since last purchase can imply higher probability of purchase, and a positive number suggests the opposite. The median of fixed effects of Recency is -2.2, which generates an odds ratio of 0.11, meaning that the customer who purchased most recently is 9 times more likely to buy the product than the customers who have not purchased the product for the longest time. And for most classes of products, the fixed effects are negative and significantly different from zero, which indicates that a more recent purchase leads to a higher purchase probability. One may argue that logically consumers would not purchase again if they have purchased products recently. However, the predicting is on a monthly scale, and for most consuming products or products that are frequently used in projects, being repurchased monthly is reasonable. Also, for those 12 classes with positive estimated fixed effects, none of the effects are significantly different from zero. Therefore, we conclude that for most classes, if a product is purchased more recently, it is more likely to be purchased again within months.
Figure 8: Distribution of estimated fixed effect for Recency

The histogram for fixed effects of Frequency shows a mode around 10 and spreads out to the right, which means that purchasing more frequently indicates higher probability of purchase. Higher estimated coefficient for the normalized frequency means that the future purchasing possibility is strongly related to the past purchasing frequency. The estimated parameters spread out from 0 to 100 because the importance of past frequencies of purchasing depends largely on the type of the product. None of the six classes with negative estimated parameter for frequencies is significant with 95% confidence level. With all significant frequency
efficient positive, we could infer that the odd ratio is greater than 1, thus we conclude that for most classes, if a product is purchased more frequently, it is more likely to be purchased again within months.

Figure 9: Distribution of estimated fixed effect for Frequency
The histogram for fixed effects of Monetary (Figure 10) shows that majority of the categories have a positive fixed effect for M which implies that spending more money in the past indicates larger probability of repurchasing. However, most of the estimated fixed effects for M are not significantly different from 0. Thus we cannot make a conclusion that more money spent implies more purchase probability. This insignificant could be explained by the large variance of money across different classes, products, and even brands. Or simply, monetary is not a good predictor of consumer’s future purchase. Also, those categories showing significance fixed effect of M spreads out from negative to positive, which made it difficult to find the patterns of the relationship between categories and the estimated fixed effect for Money. We also plot the histogram variance component from the random effect of money for each class in Figure 11 and 12. Most of the variance component of M are quite small (smaller than 20%), only some of them are large, in other words, less than 20% of the variance of total variance of M is attributable to households. Thus we may conclude that for most of the categories, the effect of spent money on probability of purchasing does not vary much across households.
Figure 10: Distribution of estimated fixed effect for Monetary
Figure 11: Variance component from the random effect of monetary for each product class
Even though Monetary is not shown to be significant for most cases, it remains as a parameter in the model since it is one of the Research Objectives of the research.

Furthermore, there could exist multicollinearity or interactions in the model, which means two or more predictors in the model are correlated. For example, if the interaction effects are considered, and we may find significant effects of interaction Monetary*Frequency on purchase probability. In this case, leave Monetary predictor in the model is necessary despite
that fact that it is not significant by itself. More advanced prediction models with interaction terms could be studied in the future.

Next, to check our model performance in predicting purchase probability, overall ROC (Receiver Operating Characteristic) is plotted based on testing data in Figure 11, which show that GLMM model achieve a good predict accuracy with AUC (Area Under the Curve) equals to 0.813, which has more accuracy predictive power than basic logistic modeling that generates AUC as 0.78. Although there is not much higher improvement for GLMM than baseline logistic model, GLMM is more interpretable and more sense logical since the scales of RFM varies across classes and assuming equality of parameters across classes renders the generalization of the research. Using GLMM model with RFM features, this study achieves a good prediction in purchase probabilities to help generating targeted promotions through sending each customer his top items promotions with highest predicted purchase probability (customer driven promotions) or sending certain products to customers with highest purchase probabilities (products driven promotions).
Figure 13: ROC Curve for GLMM and Baseline Logistic Regression
A. Research Objectives 1-8 discussion and conclusion

The objective of this study is to build a touch point based customer experience model and examined the relationships between customer’s overall shopping experience and customer’s experience on seven touch points that are proposed by Stein and Ramaseshan’s (2016) qualitative study. Also, the influence of consumer’s shopping channel choice on these relationships is examined through multi-group analysis.

The results of the study (Figure 14) show that atmospheric touch point experience has a significant positive influence on overall customer experience. This finding is consist with the past research that suggest that both in store and online atmosphere influence customer emotion, loyalty, satisfaction, and experiences. (Yang & Young, 2009; Bagdare, 2014; Fakharyan, Omidvar, Khodadadian, Jalilvand, & Nasrolahi Vosta, 2014; Mishra, Sinha, & Koul, 2014).

Communicative touch point experience is found to have a significant positive influence on overall customer experience, which confirms Khan and Rahman’s (2016) research that mass media impression positively influence brand experience.

The research also reveals a positive significant impact of process touch point experience on overall customer experience. This result support the argument that retail service process convenience is positively related to customer satisfaction and experience (Seiders et al., 2000,
Further, Employee-Customer interaction experience is found to positively influence overall customer experience, which verifies the past findings that suggest point-of-sales assistant positively influence retail brand experience (Khan & Rahman, 2016).

The positive impact of product touch point experience on overall customer experience is also supported, which is in line with the conclusion that consumer’s perceived quality and price positively influence customer satisfaction and brand loyalty. (Lu & Xu, 2015; Frank, Herbas Torrico, Enkawa, & Schvaneveldt, 2014, Dapkevicius & Melnikas, 2009).

However, technological touch point experience shows a non-significant influence on overall customer experience even though previous research points out that shopping technology such as customization has a direct influence on positive consumer experience (Yang & Young, 2009). This inconsistency may due to the fact that technologies are the fundamental elements of each touch point, and customers may not recognize technologies as certain experience. For example, customer may consider cutting-edge digital fitting room as part of the store atmospheric experience rather than an independent technology experience.

Customer-Customer touch point experience also displays a non-significant influence on overall customer experience. This contradicts to the previous finding that customer-to-customer interaction is an important element of shopping experience and has a positive indirect influence on customer satisfaction (Parker & Ward, 2000; Harris et al., 1997). This dilemma could be due to many factors. First of all, majority of the respondents in this study age from 18 to 30, who are also be considered as Generation Y or Millennial (born from
early 1980s to early 2000s). Even though Millennial Generation are preoccupied by the way they are perceived by peers and groups (Parment, 2013), and reference group and social media shape and moderate their whole decision making process (Turley, 2014), Millennial Generation are also very desired to express their uniqueness (Talbott, 2012). Thus, a true level of consumer-customer touch point experience may not be presented through survey format since the respondents may simply do not want to admit their choice being influenced by others’ opinion or advice during the survey. This explanation may be further proved by the fact that the mean of customer-customer touch point is 3.38, which is the second lowest among eight latent variables (the lowest mean is technological touch point). The lower the value, the larger the extent they disagree with the items.

Second, since Millennial Generation are powered with technologies, and they are constantly checking out reviews, when they may not see the online reviews as a format of customer-customer interactions, but merely online sources. Third, Bruhn, Schnebelen, and Schäfer (2014) concluded that the quality of customer-to-customer interactions in B2B brand communities has a positive impact on functional, experiential, and symbolic brand community benefits, which in turn influences brand loyalty. Thus, customer-customer interaction may only provide an indirect effect on overall customer experience, which could be investigated in future study.
Figure 14: Touch points based omni-channel experience results for all customers

(Red line indicates non-significant impact, Green line indicates significant impact)
Multi-group analysis is performed to investigate the influence of consumers’ channel choice (store, online, mobile, multi-channel) on the relationship between touch points and omni-channel consumer experience. The results (Figure 15) suggest that atmospheric, process, and products touch points experience are positively influence overall customer experience despite the consumer’s shopping channel (single channel or multi channel). And technological and customer-customer touch points experience are not found to significantly influence overall customer experience.

We also find that communicative touch point experience significantly influences overall customer experience for multi channel group, however, the influence vanishes when it comes to single channel group. This may because of the fact that multi-channel consumers are constantly searching deals and comparing price through different channels, when they will gain more communicative experience than single channel consumers who have less interactions with retailers during communicative.

Further, employ-customer touch point experience is found to positive influence overall customer experience for single channel consumers, but no such relationship is found for multi channel consumers. Similarly, as discussed previously, this may due to the phenomenon that single channel customers especially in-store consumers have more interaction with employees than multi channel customers do.
B. Research Objective 9 discussion and conclusion

Research Objective 9a discussion and conclusion

The significant marginal increase in purchase probability for targeted emails compared to overall emails in general could indicate that this Fortune 500 Specialty Retailer’s direct email campaign performs successfully in attracting consumers to more likely make a purchase.

However, as mentioned before, there are only about 1700 out of over 3.4 million emails falling in this type. Two possibilities could cause this sparse problem: either this retailer is unaware of the effectiveness of targeted emails, or they struggle to find more advanced
methods other than association rules (their current targeted email algorithm) to generate targeted products to suggest to consumers.

The retailer should increase the number of targeted emails to increase consumers purchase probability if the former assumption is true. In case of that the retailer needs more advanced models to predict consumers’ probability to generate targeted promotions, this study proposes GLMM method with RFM consumer behavior characteristics to project consumers purchase probability, which is illustrated in objective 9b.

**Research Objective 9b discussion and conclusion**

Through the combination of generalized linear mixed model and RFM model, we predict the probability of a customer purchasing products from a class at a specific time point. The results suggest that GLMM model with RFM variables provides a good model fit and achieve relative good predicting accuracy. Specifically, R and F are good significant predictors of consumers’ future purchase probability, while M show insignificant impact on purchase probability in most cases us not a good indicator of purchase probability. Also, by setting the household random effects, we find that effects of money on consumers’ purchase behavior do not vary much across household or consumers.
CHAPTER 6

Implication, Limitation, and Future Study

Research Objectives 1-8 Implication

Theoretical Implications

As an extension of Stein and Ramaseshan’s (2016) qualitative study, this research quantitatively examined seven consumer touch points’ experience on overall customer experience. The results confirmed the previous study on the positive impact of atmospheric, communicative, process, products, and employee-customer experiences on overall customer experience. Also, technological and customer-customer touch points failed to establish a positive relationship with overall customer experience. When examining the influence of consumer’s channel choice, this research contribute to omni-channel research by finding that communicative touch point and employee-customer touch point differ in the impact on overall customer experience between single channel customers and multichannel customers. In summary, this research provided and verified another points of view for consumer experience research.
**Practical Implications**

First of all, retailers could benefit from this research through establishing a customer touch point based customer experience strategies, and engaging today’s omni-channel consumers at these touch points during their shopping journey. Second, this research find an inconsistent impact of the touch points on overall customer experience, and previous research has claimed that multi-channel consumers do not separate channels during shopping (Blázquez, 2014). Thus, it is recommend for retailers to deliver a consistent touch points experience across different shopping channels. This includes synchronizing communicative information such as discount across all channels and providing consistent service across all channels. Last, out of expectation, this study did not find a positive influence of technological touch point experience on overall customer experience, which could be caused by consumer’s unawareness of technological touch point. In fact, research has pointed out that consumers consider mobile phones as most important source of in-store technologies (Blázquez, 2014). Thus, the adoption of new technologies across all channels may increase consumers awareness, which could in turn enhances overall customer experience.

**Research Objective 9 Implication**

**Theoretical Implications**

First of all, this study confirms the influence of RFM on prediction consumers’ behavior and values in previous RFM research (Elsner et al. 2003). Recency and Frequency are found not
only good indicates of customer values, but also good indicators to predict customers future purchase probability, i.e., the more recent and frequent the purchase is, the more likely the repurchase will happen. Monetary factor, however, is found to have insignificant influence on purchase probability. Thus this study sheds lights on the studying the role of household incomes and total money spent on consumers’ future behavior. Since many retailers segment customers merely based on total money spent, researchers could segment the consumers into groups by Monetary factors, and study and build predictive models based on these different segmentations.

**Practical Implications**

Research Objective 9 first confirms the effectiveness of targeted promotions then proposed a reasonable and good fit model to predict consumers purchase probabilities.

We believe that by sending out weekly targeted emails that are based on product association rules as well as model-based future purchase probabilities, the actual probability of the customer making a purchase will increase at least by about 14.3% as a result of this email (7.9% on the day of the email and an additional 6.4% in the week following) for each email. Considering the average frequency of the consumers of this Fortune 500 Specialty Retailer is 2 weeks per basket, the above lift will generate an additional 3.7 (0.143*52/2) purchases per customer over the course of the year. For a total 15000 email-registered customers and average revenue of $25 per customer in the dataset, this would yield an additional 1.39 million revenue per year.
Research Objectives 1-8 Limitation and Future Study

The first limitation of this study comes form the convenience student sample collected from a large university. Also, the survey questions are related to apparel and footwear industry. In order to enhance the generalizability, future research could be performed in other industries with larger random samples.

Last, Verhoef et al. (2009) suggests that different retail elements have various effects on consumers’ decision making, and consumer experiences should be viewed as the evolution of total experience over time. Thus, future study efforts could be made to conduct longitudinal research and investigate the relationship between retail drivers on consumers’ experience at different touch points as well as overall experiences.

Research Objective 9 Limitation and Future Study

First of all, this study examines the influence of Recency, Frequency, and Monetary on consumers purchase probability and builds models that are purely based on past transaction data. Other factors could also contribute to the purchasing behavior of consumers, such as geographic information, market campaigns, or even weather, product variation, customer lifestyle, and so on (Fitzpatrick 2001). Thus future study could build predictive models with comprehensives variable to predict consumers purchase probability and generate more effective targeted promotions.
Second, the observational nature of our data makes it difficult to make assuring statements about the effect of increasing emails, especially when it comes to quantifying uncertainty.

One must be cautious when making any statement that implies causation based observational data. Even though the analysis of the data puts the model in the context of increasing the revenue, we would be reluctant not include the word cautious, and the nature of this analysis renders any model-based quantification of uncertainty practically risky. In spite of this, the scale of our estimates can still provide useful information to the retailer. Future study could run a pilot test or small sample A/B test to examine the more advanced predictive models in practical world.
REFERENCES


Barga, R., Fontama, V., &Tok, W. Predictive Analytics with Microsoft Azure Machine Learning.


Fakharyan, M., Omidvar, S., Khodadadian, M., Jalilvand, M., & Vosta, L. (2014). examining the effect of customer-to-customer interactions on satisfaction, loyalty, and word-of-mouth behaviors in the hospitality industry: The mediating role of personal interaction quality and


Gatautis, R., &Kazakevičiūtė, A. (2012). CONSUMER BEHAVIOR IN ONLINE SOCIAL NETWORKS: REVIEW AND FUTURE RESEARCH DIRECTIONS. Ecoman, 17(4). http://dx.doi.org/10.5755/j01.em.17.4.3014


H.M. Chuang and C.C. Shen, A study on the application of data mining techniques to enhance customer lifetime value based on the department store industry, The seventh international conference on machine learning and cybernetics 168-173.


Philipp “Phil” Klaus, and Stan Maklan, (2012)."EXQ: a multiple-item scale for assessing service experience", Journal of Service Management, Vol. 23 Iss 1 pp. 5 - 33


Teknomo, K. (2012). ANALYTIC HIERARCHY PROCESS (AHP) TUTORIAL. Retrieved from https://pdfs.semanticscholar.org/7e27/b5a124c2e6829e1ff0d3e1279c2dbc9ebe2a.pdf


APPENDICES
Appendix A. IRB Consent Form

North Carolina State University
Institutional Review Board For The Use of Human Subjects in Research

GUIDELINES FOR PREPARATION OF INFORMED CONSENT FORM

PLEASE READ ALL OF THIS INFORMATION CAREFULLY PRIOR TO COMPLETING THE CONSENT FORM

An Informed Consent Statement has two purposes: (1) to provide adequate information to potential research subjects to make an informed choice as to their participation in a study, and (2) to document their decision to participate. In order to make an informed choice, potential subjects must understand the study, how they are involved in the study, what sort of risks it poses to them and who they can contact if a problem arises (see informed consent checklist for a full listing of required elements of consent). Please note that the language used to describe these factors must be understandable to all potential subjects, which typically means an eighth grade reading level. The informed consent form is to be read and signed by each subject who participates in the study before they begin participation in the study. A duplicate copy is to be provided to each subject.

If subjects are minors (i.e. any subject under the age of 18) use the following guidelines for obtaining consent:

- **0-5 years old** – requires signature of parent(s)/guardian/legal representative
- **6 – 10 years old** - requires signature of parent(s)/guardian/legal representative and verbal assent from the minor. In this case a minor assent script should be prepared and submitted along with a parental consent form.
- **11 - 17 years old** - requires signature of both minor and parent/guardian/legal representative

If the subject or legal representative is unable to read and/or understand the written consent form, it must be verbally presented in an understandable manner and witnessed (with signature of witness). If there is a good chance that your intended subjects will not be able to read and/or understand a written consent form, please contact the IRB office 919-515-4514 for further instructions.

*For your convenience, attached find a sample consent form template that contains necessary information. In generating a form for a specific project, the principal investigator should complete the underlined areas of the form and replicate all of the text that is not underlined, except for the compensation section where you should select the appropriate text to be used out of several different scenarios.*
North Carolina State University
INFORMED CONSENT FORM for RESEARCH
Omni-channel Consumers Journey Intervention Strategies

Principal Investigator
Jinzhao Lu

Faculty Sponsor (if applicable)
Trevor J. Little

What are some general things you should know about research studies?
You are being asked to take part in a PhD dissertation research study. Your participation in this study is voluntary. You have the right to be a part of this study, to choose not to participate or to stop participating at any time without penalty. The purpose of research studies is to gain a better understanding of a certain topic or issue. There is no personal benefit from being in a study. There is no risk posed to those that participate. In this consent form you will find specific details about the research in which you are being asked to participate. If you do not understand something in this form it is your right to ask the researcher for clarification or more information. A copy of this consent form will be provided to you. If at any time you have questions about your participation, do not hesitate to contact the researcher(s) named above.

What is the purpose of this study?
The purpose of the study is to study the consumer behavior in today’s omni-channel retailing era. It is important because it helps retailers build a more user-friendly online and offline shopping platform that helps improving consumer experience.

What will happen if you take part in the study?
If you agree to participate in this study, you will be asked to fill a questionnaire, which will take you about 8-10 minutes. No name will be asked in the survey. No identifiers will be disclosed.

Risks
There is no risk involved in the survey.
Benefits
There will be no direct or indirect benefits to the participants.

Confidentiality
The information in the study records will be kept confidential to the full extent allowed by law. Data will be stored securely in the password secured PI’s computer at North Carolina State University. No reference will be made in oral or written reports, which could link you to the study. You will NOT be asked to write your name on any study materials so that no one can match your identity to the answers that you provide. (Only use this statement if it is true for your study).

Compensation
For participating in this study “You will not receive anything for participating.”

What if you have questions about this study?
If you have questions at any time about the study or the procedures, you may contact the researcher, Jinzhao Lu at College of Textile, North Carolina State University, or [919-518-3876].

What if you have questions about your rights as a research participant?
If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Deb Paxton, Regulatory Compliance Administrator, Box 7514, NCSU Campus (919/515-4514).

Consent To Participate
“I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may choose not to participate or to stop participating at any time without penalty or loss of benefits to which I am otherwise entitled.”

Subject’s signature_________________________ Date _______________
Investigator's signature_______________________ Date _______________
Appendix B. Questionnaire

Omni-Channel Customer Experience Survey:

Thank you for participating in this survey from North Carolina State University. It will take you about 10 minutes to finish the questionnaire. Through this survey, we are interested in investigating customer shopping experience in today’s omni-channel era. YOUR PARTICIPATION IS COMPLETELY VOLUNTARY. You are free not to participate or stop participation any time before you submit your answers. If you have any questions about this survey, please contact Jinzhao Lu by email: jlu17@ncsu.edu.

Section A

1. Think and write down the retailer or brand you buy apparel or footwear from: ________________

2. Click the approach you are using during the shopping process (Select only if you use it frequently)
   ● In store
   ● On computer
   ● On mobile devices (Smart phones, tablet, etc)
   ● Others ______

On the next two pages, there are statements about your evaluation of this retailer/brand. Please indicate the extent to which you agree or disagree with each statement by clicking a number between 1 and 5.
<table>
<thead>
<tr>
<th>Atmospheric experience of this retailer/brand (store atmosphere, web design, mobile apps design)</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. They provide comfortable environment that puts me in a better mood</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. Their shopping environment is visually appealing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. Their shopping environment increase my duration</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. Their shopping atmosphere impresses me</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communicative experience of this retailer/brand (advertising and promotion)</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I feel positive toward their advertising and promotions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. They did a good job in advertising and promotions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. Their advertising and promotions are informative</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. I react favorably to their advertising and promotions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product experience of this retailer/brand</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I'm satisfied with the quality of their products</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. I'm satisfied with their product assortment</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. According to what I pay, they provide me with good value</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. Overall, I'm satisfied with the products I bought</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Technological experience of this retailer/brand</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------------------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td>---------------</td>
</tr>
<tr>
<td>1. Their technology makes shopping convenient</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. They have technology that is easy to use</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. Their technologies are interactive</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. They have better technologies compared to other retailers/brands</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process experience of this retailer/brand</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Their shopping platform is convenient</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. I can easily find information or products needed</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. Their service processing (shipping/returning) is convenient</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. Their service processing time (shipping/returning) is short</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employee-customer experience of this retailer/brand</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Their employees understand my needs</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. Their employees solve my needs efficiently</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. Their employees put themselves in the customers’ place</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. Overall, their employees deliver excellent service</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
### Customer-customer experience of this retailer/brand

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Other customers’ opinion affected my opinion toward this retailer/brand</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. I actively sought advice/reviews before shopping at this retailer/brand</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. I observed where my friends were shopping before selecting this retailer/brand</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. Other consumers positively review about this retailer/brand</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### General omni-channel customer experience of this retailer/brand

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am happy with the product I purchased from this retailer/brand</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. I am happy with the service they provide</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. My choice of this retailer/brand meets my expectations.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. Overall, my experience with this retailer/brand is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### Section B

1. What is your gender?
   - Male ( )
   - Female ( )

2. What is your age?
   - 18 – 20 years old ( )
   - 21 – 23 years old ( )
   - 24 – 26 years old ( )
   - 27 – 30 years old ( )
   - >30 years old ( )

3. Education
   - Freshman ( )
   - Sophomore ( )
   - Junior ( )
   - Senior ( )
   - Graduate ( )
   - Other (please specify)__________
Appendix C. Survey Data Snippet

<table>
<thead>
<tr>
<th>Q3</th>
<th>Q4_1</th>
<th>Q4_2</th>
<th>Q4_3</th>
<th>Q4_4</th>
<th>Q8_1</th>
<th>Q8_2</th>
<th>Q8_3</th>
<th>Q8_5</th>
<th>Q18</th>
<th>Q19</th>
<th>Q20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Think and write down a brand (retailer)...</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J. Crew</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Calvin Klein</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>D&amp;G</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Free People</td>
<td>NaN</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Kenzo</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Walmart</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Nike</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A&amp;F</td>
<td>1</td>
<td>NaN</td>
<td>1</td>
<td>NaN</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q3</th>
<th>Q4_1</th>
<th>Q4_2</th>
<th>Q4_3</th>
<th>Q4_4</th>
<th>Q8_1</th>
<th>Q8_2</th>
<th>Q8_3</th>
<th>Q8_5</th>
<th>Q18</th>
<th>Q19</th>
<th>Q20</th>
</tr>
</thead>
<tbody>
<tr>
<td>296 gap</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>297 h&amp;m</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>298 levis</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>299 vs</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>300 Patagonia</td>
<td>NaN</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>301 Hollister</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>302 NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q3</th>
<th>Q4_1</th>
<th>Q4_2</th>
<th>Q4_3</th>
<th>Q4_4</th>
<th>Q8_1</th>
<th>Q8_2</th>
<th>Q8_3</th>
<th>Q8_5</th>
<th>Q18</th>
<th>Q19</th>
<th>Q20</th>
</tr>
</thead>
<tbody>
<tr>
<td>346 GAP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>347 oldnavy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>348 adidas</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Appendix D: Wharton Data Proposal Submit Link

http://wcai.wharton.upenn.edu/research/using-purchase-history-to-identify-customer-projects/