The Impact of Experiential Store on Customer Purchases

Khaled Boughanmi, Raghuram Iyengar, and Young-Hoon Park*

January 20, 2024

Abstract

Despite the rising popularity of non-traditional retail spaces providing immersive experiences, empirical evidence on their impact on customer behavior remains limited. We study the causal impact of customers visiting an experiential store on their purchase behavior. Analyzing individual-level transactions from a personal care business over a year before and after the store's launch, we find a positive and economically significant average treatment effect on customer spending. However, substantial heterogeneity exists, with only around 20% of customers exhibiting a significant positive effect, while the majority show no significant change. The most substantial treatment effects are observed among high-value customers who, despite a long lapse since their last interaction, actively engaged with the firm. We decompose the treatment effect across the differing needs using a model that links product purchases in a customer basket to underlying customer needs. We find that needs linked to sophisticated skincare routines, connecting to high-priced items that customers can assess through hands-on testing and workshops provided in the store, exhibit positive significant effects. In contrast, treatment effects associated with basic skincare routines show no significant impact. The results align with experiential learning and haptics, offering insights into the implications for experiential retailing.

Keywords: Experiential retail, Retailing, Customer engagement, Causal inference, Probabilistic machine learning, Generalized random forest, Topic models, Customer needs

*Khaled Boughanmi is Assistant Professor of Marketing at the Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853; email: kb746@cornell.edu. Raghuram Iyengar is Miers-Busch W'1885 Professor and Professor of Marketing at the Department of Marketing at the Wharton School, University of Pennsylvania, Philadelphia, PA 19104; email: riyengar@wharton,upenn.edu. Young-Hoon Park is the Sung-Whan Suh Professor of Management and Professor of Marketing at the Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853; email: yp34@cornell.edu. All authors contributed equally and are listed in alphabetical order. We thank several managers at the participating company for their efforts in making this research possible.

INTRODUCTION

Effective customer engagement is one of the top priorities for businesses (Forbes 2021). While advances in digital technology have helped firms to achieve these objectives, many are currently facing diminishing returns from investments in digital marketing due to heightened competition (e.g., Kahn et al. 2018; Kannan and Li 2017; Lemon and Verhoef 2016). Consequently, physical stores are reclaiming their importance in shaping the customer journey (e.g., Cui et al. 2021; Grewal et al. 2009; Zhang et al. 2022).

In the retail landscape, firms have employed various strategies with their physical storefronts. Some retailers (e.g., Walmart, Costco) emphasize extensive inventories and competitive pricing, catering to a wide customer base. In contrast, others (e.g., Hermes) prioritize exclusivity and luxury goods, often showcasing their latest offerings in flagship stores located within prime real estate areas. Amidst these extremes, numerous mainstream brands are exploring new ways to distinguish themselves, influenced by consumer expectations regarding the role of brick-and-mortar stores. Recent surveys indicate that over a third of consumers seek experiential engagement with brands and prefer interacting with products before making their purchase decisions.[1]

To stay relevant and foster deeper relationships with customers, many retailers are establishing innovative stores that extend beyond conventional retail transactions. These non-traditional retail spaces, often termed as "experiential retail stores" or "experiential stores," focus more on educating and entertaining customers through immersive and interactive experiences and less so on sales (e.g., Jahn et al. 2018). For instance, Samsung 837, Samsung's experiential retail space in New York City, invites visitors to immerse themselves in exploring and interacting with Samsung's latest technology. Several brands across different industries (e.g., Nike, Glossier, Lululemon, Dyson) are also opening such stores to address their customers' needs around "shoppertainment" as well as to educate them about their offerings.

Firms adopting experiential retail strategies aim to achieve a diverse set of objectives. These may include boosting store traffic, fostering social engagement, educating customers, and gathering feedback from them. The ultimate measure of success in these endeavors often hinges on the extent to which these immersive physical spaces meant for educating and entertaining customers also function as a gateway for incremental customer purchases (e.g., Breugelmans et al. 2023). For the latter, concerns arise from past work on customer education and the implications associated with assisting customers in building their expertise. For instance, customers may develop sufficient knowledge about products through visits to an experiential store, potentially making it easier for them to switch to competitors (e.g., Bell et al. 2017; Levitt 1980). However, past work on ongoing search suggests that customers acquire knowledge outside the usual purchase journey, which can make future buying more efficient (e.g., Bloch et al. 1986). Given these diverging views, researchers have noted a growing need to examine the impact of new store formats on key performance metrics (Breugelmans et al. 2023).

Our research draws on this call to action and we assess the causal impact of customers visiting an experiential store (the treatment) on their subsequent behavior. Specifically, we are interested in addressing the following questions: Does a visit to the experiential store contribute to increased revenues for a firm? If so, how do these effects vary across customers? Moreover, what are the underlying drivers and whether store visits benefit certain product categories more than others? We address these questions in close collaboration with an Asian retailer in the personal care industry which launched an experiential store in 2019. Like many other retailers, the company was keen on experimenting with a non-traditional retail format but was uncertain about the downstream impact on customer behavior. Customers could participate in hands-on testings, interactive displays, and workshops, and notably purchases

[1] https://www.shopify.com/plus/commerce-trends/retail

were not directly facilitated within the premises. The retailer aimed to showcase its wide range of products and raise awareness in a purchase-free environment that encourages exploration of different brands and products. Leveraging individual-level transaction data spanning 24 months including both pre-launch and post-launch periods of the store, we address the first two questions using propensity score matching with the difference-in-differences procedure (Angrist and Pischke 2008) and estimate the treatment effect on customers who visited the store. Additionally, we obtain individual treatment effects by applying generalized random forests (Athey et al. 2019).

The findings show that a visit to the experiential store is effective in lifting customer purchases. On average, treated customers increase their purchase amount by about $6.00 per month over a period of 12 months post treatment, compared with a group of control customers. The effect is economically significant when considering the average monthly spending is about $20 during the pre-treatment period. Additionally, treated customers make more frequent purchases and purchase more items. Our findings are robust to self-selection, different treated and control groups, model mis-specification and different outcomes of purchase behavior. More relevant from a managerial perspective, there is substantial variation in the treatment effect across customers. Specifically, the change in purchase amount post treatment ranges from below $-2.00 to over $50, exhibiting a highly skewed distribution. The median change in purchase amount stands at just less than $2.00 per month. There is a small group of visitors that have significant positive treatment effects (about 20%), while a considerable portion of customers shows no significant change at all. The experiential store has a larger impact on older, high-value customers (based on past purchase patterns) who actively engaged with the firm but had a longer lapse since their last interaction. In sum, the experiential store is not equally effective for every customer.

We offer evidence for the drivers of the documented effects. We do so by proposing a needs-based modeling framework that links product purchases within a customer basket to their latent needs. Importantly, we establish a connection between the two analyses, i.e., the impact of store visits and the inferred customer needs, and decompose the overall treatment effect across the differing needs. The results indicate that there is significant variation in the treatment effect across needs. Customer needs exhibiting a significant positive impact from the treatment, such as anti-aging concerns, are associated with more elaborate skincare regimens. These advanced needs are related to high-priced premium products that customers can better assess through hands-on testing and participation in workshops, which are exactly the kinds of activities offered in the store. In contrast, treatment effects associated with needs related to basic and simple skincare routines are not significant. Thus, our findings are in line with past work on experiential learning (e.g., Kolb and Kolb 2005; Zhang et al. 2022) and haptics (e.g., Peck and Childers 2003). These findings are also largely consistent with recent work on customer education initiatives (Bell et al. 2017), which suggests that educating customers on firm-specific expertise can be beneficial, such as understanding the unique benefits of a specific anti-aging cream offered by the firm. In contrast, building market-related expertise, like gaining more knowledge about basic products, might not yield similar benefits.

Our paper is related to several streams of research. Broadly, our study adds to the literature covering the disruptive changes in the retailing industry. Kahn et al. (2018) provide an excellent discussion on topics such as omnichannel marketing, digital technology, and the integration of big data for innovative customer interactions. In the same vein, Gauri et al. (2021) discuss the evolving landscape of retail formats and speculate on the future of physical stores. They also offer a summary of work in related areas such as multi-channel marketing, assortments, and showrooming (refer to Table 1 on page 45). Relevant to our investigation, Gensler et al. (2017) find that non-price factors like perceived product quality and service wait time in physical stores provide valuable information to customers and contribute positively to showrooming. The notion that technology aids in product discovery and fit assessment aligns with the idea of shoppertainment, which provides an entertaining experience while potentially boosting sales, either directly or indirectly. Recent studies have sharpened our understanding of this concept, establishing it as a primary motivation for consumers to visit stores (e.g., Breugelmans et al. 2023; Jahn et al. 2018; Robertson et al. 2018). Our context serves as an exemplary case of a company leveraging shoppertainment to engage customers. Through product demonstrations, hands-on testings, interactive displays, and workshops, the aim is to foster customer engagement and education without immediate purchase pressure, potentially influencing future sales.

Our work also complements and extends the research exploring the impact of shoppertainment on future customer-level outcomes. Recent studies have focused on examing the causal impact of livestream shopping on future purchase (e.g., Liu 2022). However, within the domain of experiential stores, there is limited research employing a causal framework. Jahn et al. (2018) describe an experiment where participants visit an experiential flagship store in a simulated setting or a regular brand store and report that positive retail experiences can influence purchases of both standard and exclusive products. One notable study employing a causal framework using secondary data is Bell et al. (2020), which collaborated with a digital apparel company (Bonobos) to quantify the impact of customer visits to their physical locations (termed as Guideshops) on future revenues. We complement their work by examining the heterogeneity in the effects across customers and conducting a detailed analysis to uncover latent needs. In addition, we show how to decompose the treatment effect across these needs and offer a more nuanced understanding of the product categories that benefit the most from the visit. Finally, in contrast to Guideshops and similar physical spaces of digital native companies where customers can receive personalized assistance prior to making purchases in their e-commerce channels, our study's context emphasizes customer exploration, education, and engagement in a purchase-free environment.

Finally, we contribute to a growing literature in marketing that have integrated machine learning methods for causal analysis (e.g., Goli et al. 2021; Iyengar et al. 2022; Unal and Park 2023). In addition to our main analysis, we extend our findings with results from a needs-based modeling framework (e.g., Lee and Ariely 2006). This analysis is important from both theoretical and managerial perspectives, as it helps test potential drivers of our observed effects and provides managerial implications for the findings. While past work has explored a similar approach when examining market baskets for product recommendations (e.g., Jacobs et al. 2016), we contribute methodologically by decomposing the overall treatment effect, obtained from a difference-in-difference analysis, across differing needs. This approach allows us to determine how much each need contributes to the overall treatment effect and the mapping of products to different needs. It is noteworthy that we allow a product to serve multiple needs based on the other products within the basket, thus capturing the synergy among products in the purchased assortment.

The remainder of the paper is organized as follows. Section 2 describes our research setting and data. Next, in Section 3, we discuss our empirical methodology. Section 4 presents the findings from the main analysis and offers several robustness checks. We then describe a needs-based modeling framework in Section 5 that captures the underlying drivers of our results. Section 6 concludes with a discussion of limitations and directions for future research.

RESEARCH SETTING AND DATA

In response to the evolving shopping patterns and customers expecting to seamlessly engage with brands on both online and offline channels, firms are evaluating the role of physical spaces in the customer journey. On one end, smaller digital native companies are increasingly opening showrooms. These spaces offer customers with the opportunity to engage with a select range of products, and make informed decisions before being directed to finalize their purchases through an online channel (Bell et al. 2020). Examples abound, such as Bonobos, Warby Parker, and Casper. Bonobos, initially an online men's clothing brand, introduced "Guideshops," which are physical spaces that blend the ease of online shopping with personalized in-person service. A visit to a Guideshop provides a one-on-one experience with a dedicated guide who will go through the product line with customers and follow up with a summary of all the sizing, fit and style information they need to place their order.

Mainstream retailers, long accustomed to conducting business via physical stores (as well as online), have also expanded their approach by incorporating non-traditional retail spaces into their business strategy. Among the diverse array of non-traditional spaces, one prominent model is the experiential store, also referred to as experiential retail. These spaces prioritize creating immersive and engaging experiences for customers that extend far beyond conventional shopping encounters and typically there is no pressure for customers to make purchases within the store. Experiential stores often incorporate various elements such as interactive displays, hands-on product demonstrations, workshops, events, or themed environments. The goal is to evoke emotions, stimulate senses, and create a connection between the customers and the brand, fostering brand loyalty. The hope also is that if the experience is more enjoyable and memorable, it will in turn encourage consumers to purchase. Examples include Samsung 837, Nike House of Innovation, Petco and others.

Despite frequent mentions in popular media and praise from practitioners, there is little empirical research that has rigorously examined the impact of experiential store on customer behavior.[2] Moreover, as noted earlier, there are concerns based on past work on customer education programs and whether these immersive physical spaces meant for educating and entertaining customers can function as a gateway for incremental customer purchases (e.g., Breugelmans et al. 2023). In this paper, we address this gap and investigate the impact of customers visiting an experiential store on their subsequent behavior by collaborating with an Asia-based retailer. Operating through both physical and online channels, the firm specializes in a variety of personal care categories (e.g., skincare, makeup, body care) and offers an extensive range of products, including affordable entry-level options and premium products, to cater to the wide-ranging individual needs. In October 2019, the retailer introduced an experiential store, which offered a diverse range of immersive experiences designed to engage customers. For instance, customers had the opportunity to participate in a wide spectrum of activities, including product demonstrations, hands-on testings, interactive displays, and engaging workshops, among various other encounters. Importantly, customers could not make direct purchases at this store. The store's opening was communicated through mass marketing efforts and, to the best of our knowledge, there was no specific reason for the retailer choosing October as the launch month. Like many other firms in its industry, the company was keen on understanding the impact of experiential stores on future customer behavior. Managers recognized, however, that there might be considerable variation among customers and some products may experience bigger changes in sales as compared to others; they were therefore eager to facilitate a study.

Our data include 17,728 customers who visited the experiential store between October 2019 and March 2020. These customers presented their loyalty card during their visits, allowing the firm to keep a record of their visit.[3] However, our data lack information on customers'

[2] As an example, please see the following: https://www.ibm.com/blog/experiential-shopping-hybrid-retail/

[3] Our data do not include information on customers who visited the store but did not provide their loyalty cards, nor does it include customers who were not enrolled in the company's loyalty program. Upon customers' entry into the store, store managers encouraged them to present their loyalty cards. For those who did not have one, the team made an effort to facilitate their enrollment in the loyalty program. While some customers might opt not to participate in the loyalty program,

interactions with specific activities (e.g., hands-on testings, interactive displays) during their visits. Thus, our documented effects on customer behavior should be interpreted as the cumulative result of offering a mix of different experiences to customers. In the conclusion section, we note that future research may want to assess the impact of specific activities on future purchase behavior.

Our context is suitable for investigating the impact of customers visiting an experiential store on their behavior for several reasons. First, store managers were instructed to obtain visitor information through loyalty cards, which is essential in tracking data at the individual level. Absent this information, it would be challenging to conduct an individual-level analysis and control for many confounders. Second, our partner firm has a significant market share in its industry and offers a wide variety of products, spanning from entry-level to premium options across various categories. This extensive product portfolio gives a unique opportunity to explore the drivers of potential effects. Third, the company wanted to assess the impact of immersive experiences provided to customers in a purchase-free environment. Store managers actively encouraged customers to explore diverse brands and offerings. As a result, our findings represent a conservative estimate of the impact of store visits on future behavior, considering that some experiential stores also facilitate direct purchases.

The data consist of three parts: transaction data detailing customer purchase and return behavior, clickstream data capturing customer online activity, and socio-demographics. The transaction data contains comprehensive details at both the individual and product level, covering transaction timestamps, purchase amounts, as well as brand- and category-related information. Additionally, this data includes details regarding product returns. The clickstream data offers a granular view of individual-level interactions with the e-commerce website. It contains all of customer activities during their website visits, including visit times and pages viewed. Lastly, the socio-demographic information for customers includes their age, gender, and address, which allow us to further control for customer heterogeneity.

Using the transaction data, we define a set of outcome measures related to customer purchase behavior. Given the potential impact on customer behavior in both online and offline channels, these measures include transactions across both channels. We choose to construct these measures at the customer-month level, which serves as our primary unit of analysis. This approach aligns with the typical reporting practice of firms and improves the stability of the results by mitigating the inherent variability associated with daily or weekly data. Given our primary focus on assessing the impact on sales, our key metric is the monthly purchase amount spent by each customer (purchase amount).[4] Additionally, we consider other aspects of customer purchase behavior, including the number of purchases made (purchase frequency) and the number of items purchased (purchase quantity). Doing so allows us to disentangle changes in purchase amount, identifying whether changes are due to changes in frequency and/or quantity.

Finally, considering that changes in customer behavior might be influenced by shifts in the assortment or basket of products purchased by customers, we explore a few metrics to examine basket dynamics. By differentiating between new and familiar products based on whether the customer had previously purchased them in the pre-treatment period, we analyze the number of new products purchased each month (new products). Additionally, we distinguish between the highest and lowest-priced products purchased within the monthly basket at the customer level (most and least expensive, respectively). These metrics, serving as proxies of engagement with the firm, help assess purchase behavior at a granular level and are useful to examine how customer behavior changes following visits to the store.

---

and others might visit without using their card, it is reasonable to assume that the number of such customers is small.

[4] All transactions were recorded in the currency of the country where the company's headquarter was located. We converted the purchase amount to U.S. dollars using the average exchange rate over the data period.

EMPIRICAL FRAMEWORK

In this section, we outline our identification strategy and describe both the treated and control groups. We then discuss the details of our approach to estimate the effects.

Identification Strategy

Our primary objective is to examine how customer behavior changes as a result of their visit to the experiential store and explore the heterogeneity of the effect. We define the treated group as a cohort of customers who visited the store within the same time frame (e.g., the same month). Analyzing the behavior of a cohort is helpful as it allows us to identify pre-treatment and post-treatment periods clearly for making comparisons between customers who visited and those who did not, while ensuring they share similar characteristics (e.g., Goldfarb et al. 2022; Iyengar et al. 2022). To mitigate concerns with unobservable confounders, we deliberately exclude customers who visited the store in the first few months after opening, as they may exhibit systematic differences from other customers (e.g., Rogers 2003). Our main findings are based on a cohort consisting of 1,107 customers who visited the store in January 2020, about three months after the store's launch. For the purpose of comparison, we obtain a sample of 49,641 customers who had not visited with the store as of March 2021 (the end of our data period). As part of robustness checks, we replicate our analysis to include other monthly cohorts of visitors and our main findings are qualitatively similar.

Before establishing the causal impact on customer purchase behavior, we first analyze purchase amounts over a 24-month data period, excluding January 2020, for two groups: customers who visited the store only once in January 2020 (treated) and those who did not (control). In Figure 1, the first 12 months, denoted as -12 to -1, include the pre-treatment period from January 2019 to December 2019. The subsequent 12 months, labeled as 1 to 12, represent the post-treatment period, with month 0 being January 2020. As shown in Figure 1, there are significant differences in purchase patterns between these two groups. On average, customers who visited the store spend $19.81 per month post treatment, whereas those who did not visit spend $28.30. Despite this cross-sectional comparison potentially suggesting a negative value for the store visit, it is likely that customers with different pre-treatment characteristics self-select into store visits. This is evident as treated customers spend about $19.60 less per month compared to the control customers during the pre-treatment period.

Table 1 shows that treated customers differ from the control customers across multiple measures. For example, customers in the treated group are younger and reside closer to the store as compared to those in the control group. Thus, a simple comparison of customer behavior between these two groups would yield biased estimates when assessing the effect.

Propensity Score Matching

To construct a control group that is comparable to the treatment group, we implement matching by estimating propensity scores which reflect a customer's propensity of visiting the experiential store.

We estimate propensity scores through logistic regression using three sets of observed covariates from the 12-month pre-treatment period. The first set includes measures that capture the broader interaction between customers and the firm, along with their interest in its offerings. This includes pre-treatment period averages of customer purchases and returns, including: (1) time elapsed (in months) since enrolling in a loyalty program (tenure), (2) time elapsed (in months) since the last purchase (recency), (3) number of purchases made (purchase frequency), (4) number of items purchased (purchase quantity), (5) total amount spent on purchased products ($) (purchase amount), (6) number of items returned (return quantity), and

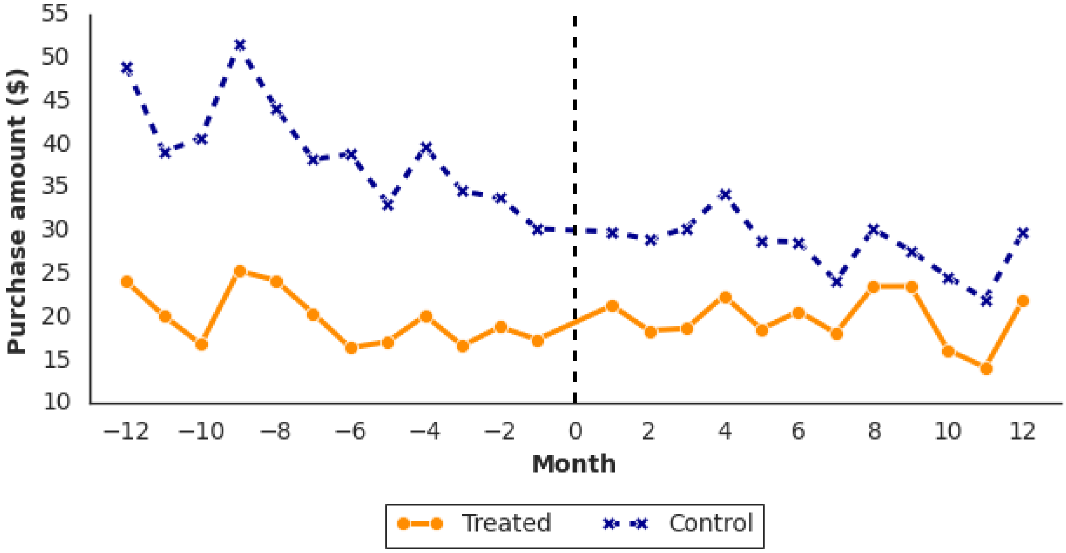


Figure 1: Purchase Amount ($) of Treated and Control Groups

Table 1: Summary Statistics of Treated and Control Groups

	Treated	Control	Difference	p-value
Customer-firm relationship				
Tenure	78.64	95.90	-17.26	0.00
Recency	4.67	3.37	1.30	0.00
Purchase frequency	0.60	0.59	0.02	0.42
Purchase quantity	2.71	2.25	0.46	0.00
Purchase amount ($)	19.83	39.43	-19.60	0.00
Return quantity	0.02	0.02	-0.01	0.15
Return amount	0.43	1.15	-0.72	0.00
Online activity				
Website visits	1.02	0.33	0.69	0.00
Page views	1.98	0.66	1.32	0.00
Socio-demographics				
Age	33.74	45.17	-11.43	0.00
Gender	0.94	0.93	0.01	0.21
Address	0.50	0.14	0.37	0.00
Distance	0.36	1.35	-1.00	0.00
Customers	1,107	49,641		

(7) total refund amount for returned products ($) (return amount). The second set includes the pre-treatment period averages of online activity measures: (8) number of visits to the website (website visits) and (9) number of product pages viewed (page views). The third set of covariates includes customer socio-demographic variables: (10) age and (11) gender (1 for female and 0 otherwise). Additionally, we include (12) address (1 if address information is available and 0 otherwise) and (13) the Manhattan distance to the store (distance). The distance is calculated using the coordinates of customers' addresses and the store's location. Both address and distance measures are included to account for unobserved socio-demographics, such as education, income, lifestyle, which could potentially influence store visits. Table 1 shows the summary statistics of the covariates which exhibit significant differences across various measures between the two groups.

Upon estimating the propensity score model, denoted by $e(x;\beta)$ and parameterized by β, we follow Imbens and Rubin (2015) and transform the score into the log-odds ratio:

$$l(x;\beta) = \ln\Big(\frac{e(x;\beta)}{1 - e(x;\beta)}\Big).$$

We assess the quality of the matching procedure in a few ways. First, we examine the distribution of the propensity scores within both the treated and control groups to assess their similarity after matching. Figure 2 shows the density of the estimated propensity scores for the treated and control groups, before and after the matching process. Before matching, the distributions share overlap but are significantly different from each other. After matching, the distributions closely resemble each other, indicating little bias in the difference in propensity scores between the groups.

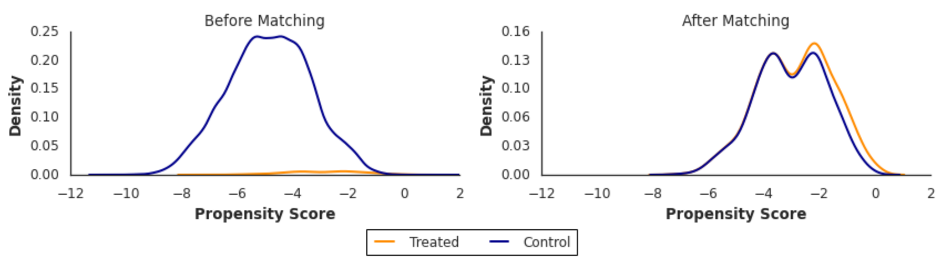


Figure 2: Distribution of Propensity Score

We also assess whether our matching procedure achieves balance in covariates between the treated and control groups. We evaluate standardized differences in the means of covariates between the two groups (e.g., Austin 2009; Imbens and Rubin 2015). Figure 3 shows the standardized differences for each variable used in propensity score estimation. The figure confirms that matching leads to a substantial improvement in covariate balance. After matching, all of the normalized differences are below 0.1, which is similar to a degree of balance observed in a fully randomized experiment (e.g., Stuart 2010).

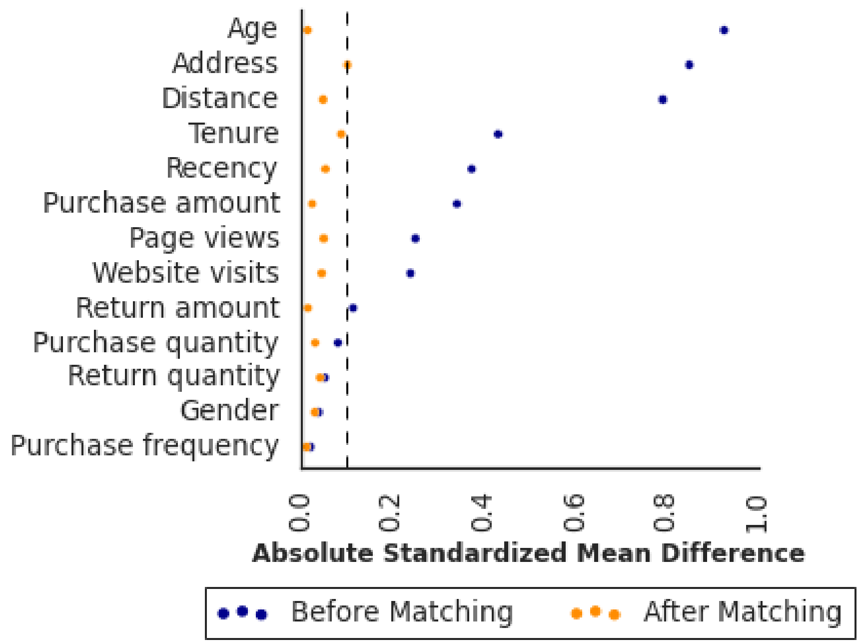


Figure 3: Covariate Balance

Difference-in-Differences

We use a sample of 1,102 pairs of treatment-control customers obtained after matching and employ a Difference-in-Differences (DD) approach to determine the impact of visiting the store on customer purchase behavior.[5] We estimate the average effects on the outcomes by employing the following DD model:

$$\mathrm{Y}_{it} = \tau \cdot \mathrm{W}_{it} + \theta_i + \lambda_t + \epsilon_{it}. \tag{1}$$

The variable Y_{it} is the outcome measure of interest for customer i in month t (e.g., purchase amount). The variable W_{it} is an indicator for treatment, equal to 1 if customer i has already visited the store by month t and 0 otherwise. The parameters θ_i and λ_t are individual- and month-fixed effects, respectively, and ϵ_{it} is the error term. Our two-way fixed-effects framework accounts for time-invariant customer characteristics, common time trends, and month-to-month fluctuations. The parameter τ measures how the store visit affects future customer behavior. To further enhance the robustness of our analysis, we employ robust standard errors clustered at the customer level to account for any potential serial correlation.

FINDINGS

In this section, we report our findings for the average treatment effects on the treated and discuss the heterogeneity of the effects.

Average Treatment Effects

Table 2 provides the treatment effects of customers visiting the store on their purchase behavior in terms of three outcomes. We observe that customers who visited the store, when

[5]Matching with replacements was implemented, resulting in 1,102 matched pairs comprised of 1,102 treated customers and 996 (unique) control customers. Consequently, we have examined data from 2,098 customers over a 24-month period, totaling 50,352 observations.

compared with their matched controls, exhibit an average increase of \$5.76 in monthly purchase amount during the 12-month post-treatment period. This increase is economically significant, especially when considering the average monthly purchase amount is approximately \$20 during the pre-treatment period.

The lift in purchase amount may be driven by an increase in purchase frequency and/or purchase quantity. Table 2 shows that treated customers, in comparison to their matched controls, exhibit an average increase of 0.10 transactions and 0.78 items per month. Compared with the pre-treatment values, these effects represent an increase of 16.7% in purchase frequency and 28.6% in purchase quantity, respectively. Thus, the experience store stimulates customer purchases by increasing the frequency of purchases among treated customers as well as making them buy a larger number of items.

Table 2: Treatment Effects Using DD

	Purchase Amount (\$)	Purchase Frequency	Purchase Quantity
Mean	5.76***	0.10***	0.78***
	(1.77)	(0.03)	(0.27)
Individual fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes
Customers	2,098	2,098	2,098
Observations	50,352	50,352	50,352

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.
Notes: Outcome measures in January 2020 (month 0) were excluded in the analysis.
Robust standard errors, in parentheses, are clustered at the customer level.

A broad analysis of the changes in product assortments within customers' baskets highlights the potential for a deeper investigation to uncover the underlying drivers of the documented effects.[6] For simplicity, we characterize a basket along two primary dimensions: its cost and the purchase of new items. For the former, we determine the highest and lowest priced products within monthly baskets. As for the latter, we categorize products as either new or familiar for each customer based on their prior purchases. Table 3 shows that treated customers, in comparison to their matched controls, purchase an average of 0.48 more new items monthly and spend about \$1.60 more per month on the most expensive product in their basket. However, we do not observe any significant impact on the least expensive product. Taken together, these results suggest that treated customers tend to buy more expensive items and new items as compared to their matched controls.

Table 3: Treatment Effects on Purchased Products Using DD

	New Products	Most Expensive (\$)	Least Expensive (\$)
Mean	0.48***	1.59***	0.39
	(0.19)	(0.43)	(0.26)
Individual fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes
No. of Customers	2,098	2,098	2,098
No. of Observations	50,352	50,352	50,352

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.
Notes: Outcome measures in January 2020 (month 0) were excluded in the analysis.
Robust standard errors, in parentheses, are clustered at the customer level.

[6]The subsequent section contains a detailed analysis of the baskets using a model that captures latent customer needs influencing their purchase patterns.

Appendix A shows the robustness of our main findings to self-selection, different treated and control groups, model mis-specification and different outcomes of purchase behavior.

Heterogeneous Treatment Effects

We explore the heterogeneity of the treatment effect by obtaining individual-level treatment-effect estimates using the Generalized Random Forest (GRF) procedure (Athey et al. 2019). GRF harnesses machine-learning principles to provide a nonparametric statistical estimation approach for causal inference in observational studies. The method is an extension of the causal forest method (Wager and Athey 2018). The fundamental concept behind GRF is to tailor the splitting criteria for constructing individual trees and identify partitions that highlight the most substantial differences in treatment effects. This feature allows us to find, through data-driven techniques, the specific features contributing to the heterogeneity in the treatment effect. Since the method is based on random forests, it can accommodate nonlinearity in relationships among features without necessitating multiple interaction terms, thereby eliminating susceptibility to the functional form in estimation. The application of individual-level treatment effects has been demonstrated in previous research across various empirical contexts, as it holds the potential to enhance personalized targeting and communication strategies (e.g., Ascarza 2018). For more details about GRF, we refer readers to (Athey et al. 2019).

Table 4 reports the treatment effects and their heterogeneity. Note that the average treatment effects are similar to our findings reported in Table 2 which are obtained using the DD approach that assumes a linear and additive treatment effect (Keele 2015). That both procedures yield average treatment effects that are comparable to one another serves as another robustness check for our model specification.

Table 4: Treatment Effects Using GRF

	Purchase Amount ($)	Purchase Frequency	Purchase Quantity
Mean	5.95***	0.10***	0.79***
	(1.77)	(0.03)	(0.28)
Minimum	-2.21	-0.11	-0.09
Median	1.94	0.06	0.36
Maximum	50.98	0.87	6.79
N	2,098	2,098	2,098
$N_{\tau \geq 0}$	1,796	1,920	2,088
$N_{\tau < 0}$	302	178	10

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: Outcome measures in January 2020 (month 0) were excluded in the analysis. Standard errors are shown in parentheses.

We find significant variation in the treatment effects across customers on all three outcomes. Table 4 and Figure 4 show the distribution of individual treatment estimates on purchase amount. A majority of treated customers (86.0%) exhibit positive effects, while the remainder (14.0%) shows negative impact. As compared to their matched controls, treated customers demonstrate an average increase of about $6.00 in monthly purchase amounts post treatment. However, the median change in purchase amount stands at $1.94, suggesting a more modest (yet still notable) impact of the store on customer purchases. Finally, 83% of customers (80% exhibiting positive estimates and all those with negative estimates) show no significant change in their purchase behavior post treatment at the 95% level. These results collectively indicate that the treatment effect on customer purchases primarily stem from a small group of customers who show significant positive effects (17%), while a considerable portion of customers shows no significant change at all.

The effects on both purchase frequency and quantity are similarly heterogeneous. These findings underscore the significance of employing the GRF procedure to obtain individual treatment effects, shedding light on the varying impact of store visit on customer behavior.

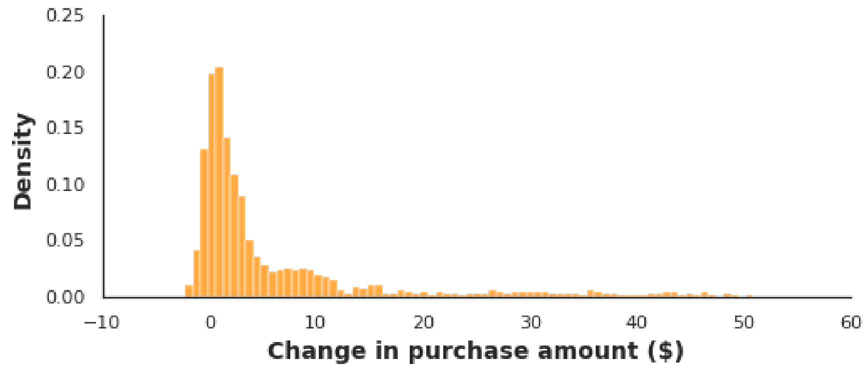


Figure 4: Distribution of the Treatment Effect

Examining the heterogeneity in the treatment effect across customers is important for both theory and practice. To that end, we leverage a useful feature in the GRF method, namely the importance measure of covariates employed in the estimation. GRF constructs its trees during training by selecting covariates that exhibit the most pronounced differences in treatment effects, thus making them crucial in understanding the source of heterogeneity. The importance measure of covariates quantifies the relative significance of each covariate in generating these splits. This measure is computed by considering the frequency with which each covariate is employed for splitting the nodes and weighting these frequencies based on the depth of each tree.

Table 5 shows the importance weight of each covariate, along with its rank among all covariates used in the analysis. The results indicate that the causal forest allocates approximately 22% of its splitting decisions to purchase amount. Moreover, measures associated with past purchase behavior, such as recency, purchase frequency and quantity, and purchase amount, collectively accounts for over 50% of the splits. This result is in line with the existing literature which highlights the significance of RFM (recency, frequency, monetary value) measures as potential moderators of various marketing activities (e.g., Rossi et al. 1996; Kumar and Shah 2004). Along these lines, tenure is important and accounts for about 8% of the splits. This variable is similar to RFM measures as it reflects the level of engagement customers have with the firm. The causal forest also spends about 6% of its splits on the age variable, suggesting the effect varies between young and old customers.

Table 5: Importance of Covariates in Heterogeneous Treatment Effects

Rank	Covariate	Importance (%)
1	Purchase amount	22
2	Purchase quantity	15
3	Purchase frequency	10
4	Gender	8
5	Tenure	8
6	Page views	7
7	Website visits	7
8	Age	6
9	Distance	4
10	Return amount	4
11	Recency	4
12	Return quantity	3
13	Address	1

Next, we relate heterogeneous treatment effects to observed covariates and categorize customers into two distinct groups based on the significance (at the 95% level) of their individual treatment-effect estimates: the significant group, comprising customers with significant treatment effects, and the non-significant group, comprising customers without significant treatment effects. Table 6 shows the results from a logistic regression that relates the individual-level covariates to the likelihood of belonging to the group that experiences a significant, positive impact on customer purchases. The results suggest that the experiential store has a significant impact on older, high-value customers (based on past purchase patterns) who actively engaged with the firm but had a longer lapse since their last interaction. These customers also have a history of higher returns in terms of quantity.

Table 6: Covariates and the Significance of Individual-level Treatment Effects

Variable	Estimate	Std. Error
Tenure	-0.01***	0.00
Recency	0.38***	0.03
Purchase frequency	1.33***	0.26
Purchase quantity	0.09**	0.04
Purchase amount	0.04***	0.01
Return quantity	7.87***	2.14
Return amount	-0.01	0.05
Website visits	-0.29	0.23
Page views	0.22*	0.13
Age	0.03***	0.01
Gender	0.56*	0.33
Address	-0.51**	0.19
Distance	0.08	0.10
Constant	-6.40***	0.53
Observations	2,098	
Pseudo-R^2	0.47	

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

EXPLANATIONS

In this section, our focus shifts towards uncovering the underlying drivers of our documented effects. To achieve this goal, we delve into a more granular analysis, leveraging the opportunity that our partner firm offers a wide variety of products across multiple categories (e.g., skincare, makeup). In our context, we observe over 15,000 distinct, non-empty baskets across customers and months, involving more than 10,000 unique products. Given the level of data richness, a formal modeling framework is necessary to extract key insights regarding underlying drivers. In what follows, we first discuss the literature relevant for the treatment effect and how the variation in the latter across categories (or products) can help us in identifying the drivers. We then propose a modeling framework that formalizes how latent customer needs are connected with product purchases and allows us to estimate the variation in the treatment effect across these differing needs.

Experiential Learning and Haptics

Experiential learning theory broadly refers to a process whereby people learn in four stages: experiencing, reflecting, thinking, and acting (Kolb and Kolb 2005).[7] In the context of acquiring product-related information, past work on sensory marketing indicates that experiential learning draws on all five senses but touch can play a key role in enhancing the remaining (Krishna 2012; Dzyabura and Jagabathula 2018; Zhang et al. 2022). Peck and Childers (2003) develop a framework demonstrating the significance of haptic information across products and retail settings. They highlight the haptic system's ability to encode material properties such as texture and sensory appeal, particularly when consumers interact directly with products instead of merely observing them from behind a retail counter. Hence, we suggest that personal care products examined in this research, which have salient material attributes contributing to customers' overall sensory experiences and influencing their perception of product effectiveness, comfort, and quality, can significantly benefit from interactions in the experiential store (e.g., hands-on testings, interactive displays, workshops). This may be less pronounced for products with with fewer distinct tactile characteristics. In the personal care industry, companies often focus on enhancing haptics to provide a more enjoyable customer experience. They create textures that feel pleasing on the skin, give a sense of hydration, or offer distinctive sensory experiences, thus influencing customer perception and satisfaction with products.

We note that the differentiation between material and non-material attributes of products aligns conceptually with the distinction between experiential and search attributes, respectively (Nelson 1974). Given this similarity, prior research offers corroboration that there is a greater significance of experiential attributes in products after direct engagement (e.g., Wright and Lynch Jr. 1995). Moreover, while multiple products may cater to the same customer need (e.g., several products collectively addressing anti-aging concerns), we hypothesize that customer needs served by products featuring salient experiential attributes benefit more from the experiential store, because experiential attributes play a more important role in influencing customer decisions within such contexts.

Decomposing the Treatment Effect

The personal care industry, our empirical context, caters to a wide spectrum of customer needs, reflecting diverse preferences, concerns, and trends. Within the skincare regimen, for in-

[7] Experiential learning also refers to a process by which consumers acquire knowledge from their previous consumption experiences (e.g., Lin et al. 2015). This is different from our framework of haptics, in which experience is characterized by a physical engagement with products.

stance, some individuals aim to meet the fundamental requirement of maintaining healthy skin by incorporating simple daily routines like cleansing, moisturizing, and sun protection. They prioritize products that fit into their everyday regimen for effectiveness. Others seek products designed to combat signs of aging, addressing concerns like wrinkles, fine lines, and skin firmness. Additionally, others prioritize hydration, opting for products that deeply replenish moisture levels in their skin. The multifaceted nature of customer needs in this domain stems from various factors. These include shifts in consumer trends, evolving health and wellness preferences, seasonal fluctuations, and ongoing technological innovations. In this research, we are agnostic regarding the potential causes behind these differing customer needs.

We assume that there exist K underlying needs that customers seek to address through their purchases (e.g., Jacobs et al. 2016; Urban and Hauser 2004). The treatment effect (τ) in Equation (1) can be expressed as follows:

$$\tau = \sum_{k=1}^{K} \tau_k, \tag{2}$$

where k denotes a latent need and τ_k represents the treatment effect related to need k.

The motivations driving a customer's purchase of a product are closely connected to their broader shopping decisions. For example, a customer might purchase a product (e.g., mask) as part of their basic skincare routines, but they may also purchase it alongside complementary items like cleansers, serums, or creams to fulfill their anti-aging needs. This implies that assuming each product belonging to a singular need is not appropriate, because a product may cater to various customer needs depending on the presence of other items in the basket (e.g., Manchanda et al. 1999; Seetharaman et al. 2005). The interaction among different products in the basket reflects the customer's comprehensive approach to meeting their needs, where each product contributes to addressing one or more requirements. Therefore, when inferring customer needs, it is critical to consider both the specific product s and the basket B_{it}, which represents all products purchased by customer i in month t (the unit of analysis in this research), highlighting the complexity of inferring customer needs in our context.

Let $\mathrm{Y}_{it,k}$ denote the outcome (e.g., purchase amount) for customer i in month t related to need k. Then, $\mathrm{Y}_{it,k}$ can be expressed as follows:

$$\mathrm{Y}_{it,k} = \begin{cases} \sum_{s \in B_{it}} p(k|s, B_{it}) \cdot \mathrm{Y}_{it,s}, & \text{if } B_{it} \neq \varnothing, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The variable $\mathrm{Y}_{it,s}$ represents the outcome related to product s purchased by customer i in month t. The term $p(k|s, B_{it})$ is the probability of having need k conditional on having product s within basket B_{it}. The symbol $\varnothing$ refers to the empty set with no purchases.

Equation (3) merits explicit mention because it is the primary focus of our efforts to uncover potential drivers for the effects. First, the term $p(k|s, B_{it})$ plays a pivotal role as it allows us to incorporate the level of uncertainty associated with the outcome related to the k-th need. In essence, within a customer's basket, this specification enables products to be associated with different needs with varying degrees. Second, the explicit consideration of the basket B_{it} is integral in establishing the link between underlying needs and products. This inclusion allows us to capture the interactions among products within a basket as they could be linked to similar needs. Third, it addresses the concern that, even if customer needs are proportionally represented in a basket, the treatment effect may differ across needs. For instance, certain needs (e.g., anti-aging) might be more expensive to address than others (e.g., simple skincare routine) and this difference is accounted in the model.

Our model conceptualizes the set of products purchased by customer i in month t (B_{it}) as assortments or bundles designed to address their needs. This product portfolio exhibits mixed membership with respect to customer needs, where each need is associated with a probability distribution across products. Each basket, in turn, is characterized by a probability distribution across customer needs. Essentially, our model integrates a mixed membership model, where monthly baskets exhibit varying degrees of association with customer needs. Moreover, these needs exhibit different levels of association with various products. As a special case, if the relationship between product s and basket B_{it} is ignored, our model collapses down to a latent class model (e.g., Kamakura and Russell 1989). In such a scenario, the treatment effect is distributed among the inferred groups, each represented by group k with its own treatment effect denoted as τ_k.

Our model accommodates two additional scenarios based on the assumed relationship between products and customer needs. First, consider a scenario where product s is directly associated with need k, facilitating the decomposition of the treatment effect across products ($\tau_k = \tau_s$). Although useful in contexts with a limited set of products, capturing the interactions among products within a basket might remain challenging under this approach. Second, suppose product s can be categorized into a single predefined group k, such as based on product categories like skincare or makeup. In this case, the term $p(k|s, B_{it}) = p(k|s) = 1$ and $p(k'|s, B_{it}) = p(k'|s) = 0$ for all $k' \neq k$. This approach involves distributing the treatment effect among these predefined groups, each represented by group k with its own treatment effect. However, this approach is evidently inadequate when the same product can satisfy a different need based on other products present in the basket.

The treatment effect corresponding to need k can be specified as follows:

$$\mathrm{Y}_{it,k} = \tau_k \cdot \mathrm{W}_{it} + \theta_{i,k} + \lambda_{t,k} + \epsilon_{it,k}, \tag{4}$$

where the parameters $\theta_{i,k}$ and $\lambda_{t,k}$ are individual- and month-fixed effects associated with need k, respectively, and $\epsilon_{it,k}$ is the error term. τ_k is the coefficient of interest, measuring the treatment effect associated with need k. Under this framework, as in Equation (2), the treatment effect τ is assured to be decomposed into the treatment effects across different needs, satisfying $\tau = \sum_k \tau_k$. We offer a proof of this decomposition in Appendix B.

The term $p(k|s, B_{it})$ stands as the most important component within our model outlined in Equations (3) and (4). To estimate $p(k|s, B_{it})$, we employ a latent Dirichlet allocation (LDA) topic model (Blei et al. 2003), a probabilistic model commonly used in natural language processing and machine learning. It is particularly powerful in uncovering latent thematic structures within a collection of documents or, in this scenario, baskets. In our analysis, LDA assumes that baskets are mixtures of latent needs or topics, with each topic representing a probability distribution across products. The model aims to uncover these needs and their distribution in each basket. It operates under the assumption that products within a basket are interrelated due to their association with shared customer needs. By applying LDA to customer- and month-level baskets, we can uncover underlying patterns of association among products within baskets and infer the latent topics that customers address through their purchases. The strength of LDA lies in its ability to identify these latent structures without prior knowledge of the topics or labeled data regarding customer needs. Because this approach makes it powerful for exploring complex data and extracting meaningful insights from unstructured information, it has been applied in previous research across a range of empirical contexts (e.g., Jacobs et al. 2016; Kim and Zhang 2023). For more details about LDA, we refer readers to (Blei et al. 2003).

In our approach, we assume that baskets are interchangeable and each basket represents a-bag-of-products. That is, the sequence of product purchases within a month is irrelevant, and the timing of product purchases within a month is considered independent. Additionally, we assume that all topics exist before the treatment and the treatment does not alter the data generating process. LDA infers the probability distributions of latent topics associated with products and the weights of these factors within baskets based on the product frequency within baskets. Since our goal is to document the treatment effect on outcomes associated with these latent topics, LDA enables the estimation of $p(k|s, B_{it}) \propto \phi_{ks}\theta_{B_{it},k}$, which we use to compute the outcomes corresponding to various topics, as defined by Equation (3). The parameter ϕ_{ks} denotes the probability of product s being associated with topic k and $\theta_{B_{it},k}$ represents the probability of topic k being associated with basket B_{it}. Appendix C provides details of the inference for the estimates $\hat{\phi}_{ks}$ and $\hat{\theta}_{B_{it},k}$.

Results

We summarize our model results across four main areas. First, we present findings regarding the identification of customer needs. Second, we examine the treatment effects specific to these needs. Third, we report the relative contribution of specific products to the treatment effect related to each need. Finally, we compare the needs-based results to those obtained when products are categorized in a predefined, exogenous manner.

We first summarize the inferences regarding latent needs in Figure 5, which presents the distribution of 20 topics associated with baskets.[8] The probability of a topic associated with a basket is calculate as follows: $\hat{\theta}_k = \frac{1}{\mathcal{B}} \sum_{B_{it} \neq \varnothing} \hat{\theta}_{B_{it},k}$, where $\mathcal{B}$ denotes the total number of non-empty monthly baskets. The topics exhibit a fairly even distribution across the baskets, indicating a wide range of customer needs impacting customer purchase behavior. This may arise from various types of customer requirements and preferences shaping their purchase decisions. For example, Topic 1 includes a broad range of products for basic skincare routines, while Topic 2 comprises specialized products aimed at addressing advanced concerns like anti-aging within the skincare domain. In contrast, Topic 11 revolves around the products using ingredients like botanical essences in masks and other related items.

[8] The LDA estimation process identified 20 distinct topics associated with customer needs. Appendix C provides details of the calibration procedure used to determine the optimal number of topics. Our data contains detailed product descriptions including brand- and category-related information. These descriptions serve to convey various aspects of products, such as their benefits, key ingredients, and their resonance with the intended audience, while aligning with the brand's image and values. For instance, product names like "Hydrating Serum" or "Vitamin C Brightening Cream" communicate the product's intended use or primary components. Additionally, certain products are delineated by their efficacy, as seen in sunscreens categorized by their Sun Protection Factor (SPF) ratings (e.g., SPF 50, SPF 30). To interpret the underlying topics, we follow Blei et al. (2003). Specifically, we identify the top ten products most closely associated with a particular topic. Subsequently, we extract and analyze the descriptions of these chosen products. Furthermore, we consulted extensively with our partner firm to understand the significance of each latent need. The partner firm validated the multitude of customer needs that our analysis identified. These needs stem from various factors, including individual preferences, specific concerns requiring tailored solutions, emerging trends, lifestyle considerations, and various occasions. Please see https://www.statista.com/statistics/1334325/most-common-skin-concerns-among-us-skin-care-shoppers-by-generation/. Due to a non-disclosure agreement with our partner firm, we are unable to provide comprehensive descriptions of these identified topics and their associated products. We instead provide a few examples to illustrate the nature of these topics.

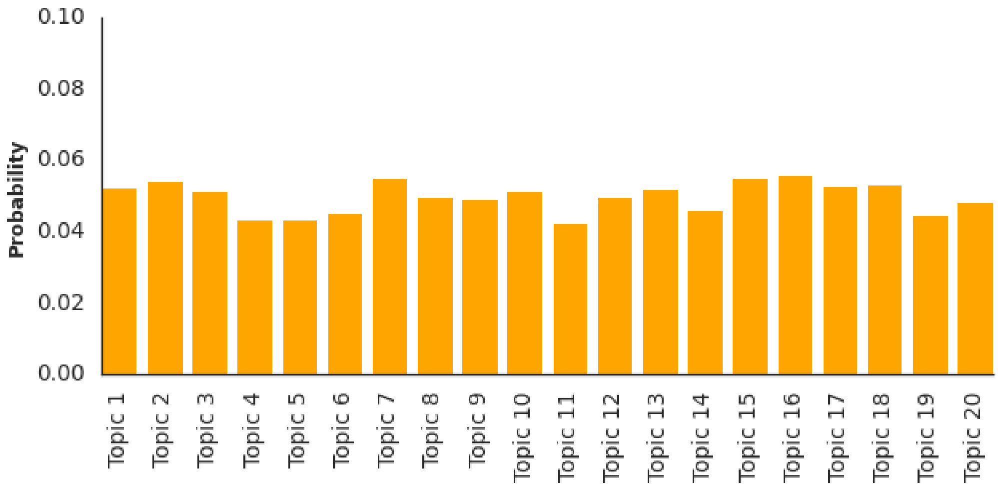


Figure 5: Distribution of Topics

Now that the topics have been identified, we examine how treatment affects various customer needs. For this purpose, we employ the DD model outlined in Equation (4). Figure 6 shows the estimated treatment effects on the purchase amount associated with each topic. The vertical line bar represents the 95% confidence intervals for each treatment effect. Note that the sum of treatment effects across all these topics totals the magnitude of the overall treatment effect reported in Table 2 ($5.76).

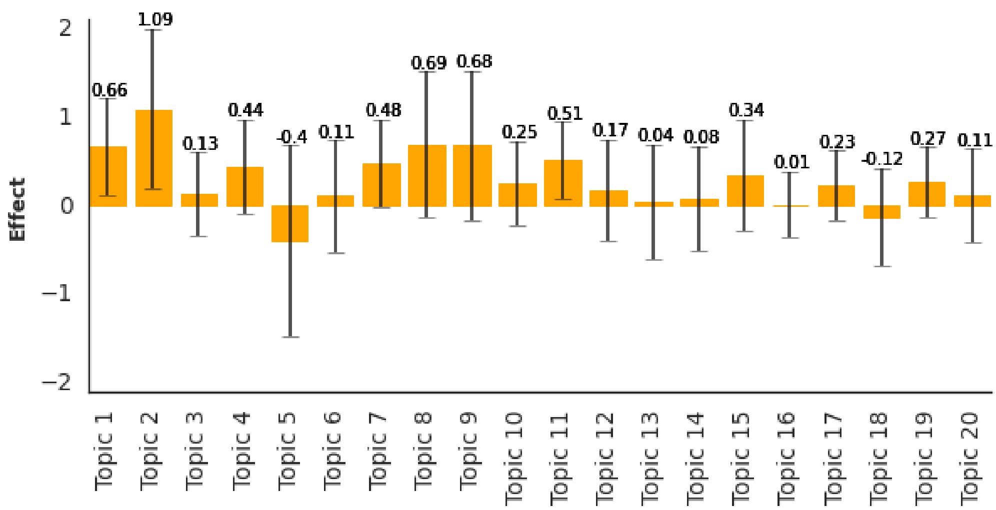
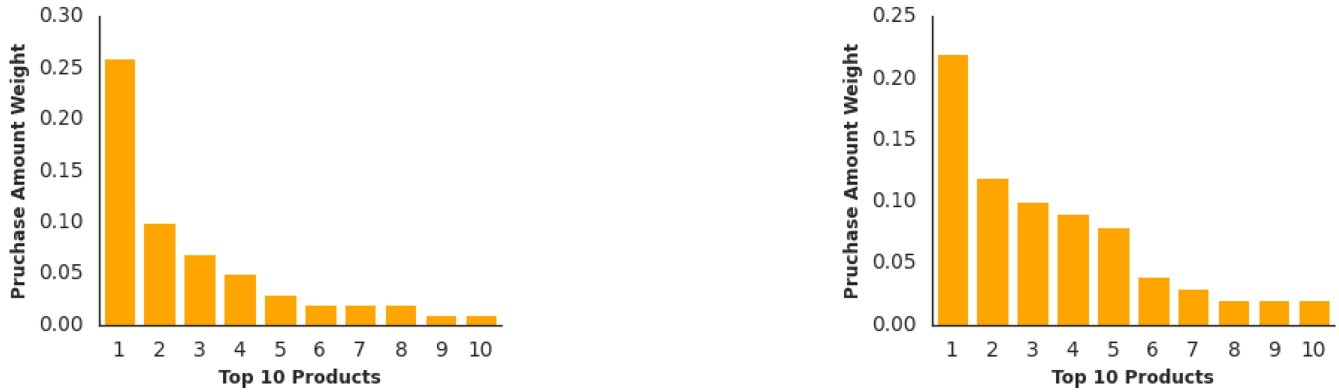


Figure 6: Treatment Effects Across Topics

Our results reveals that four topics exhibit significant effects (Topics 1, 2, 7, and 11), and are largely in line with our hypothesis based on the literature on experiential learning and haptics. For example, Topic 2, which demonstrates the most pronounced treatment effect ($\tau_2 = 1.09, p < 0.01$) contributing roughly 20% to the overall treatment effect, consists of specialized products focused on addressing advanced skincare concerns like anti-aging. These products often comprise high-priced premium items that customers can better assess through hands-on testing and participation in workshops, which are exactly the kinds of activities offered in the store. Consequently, these products significantly contribute to an increased purchase amount. Similarly, Topic 1 ($\tau_1 = 0.66, p \leq 0.05$), Topic 7 ($\tau_7 = 0.48, p \leq 0.05$), and Topic



(a) Topic 1

(b) Topic 2

Figure 7: Contribution of the Top 10 Products to Topics' Purchase Amount

11 ($\tau_{11} = 0.51, p \leq 0.01$) include a spectrum of products, ranging from cost-effective entry-level choices to luxurious premium items. Considering the complexities inherent in making purchase decisions due to numerous options available with varying benefits and price points, customers can try these products in the store and make more informed purchase decisions. Together, the four significant topics contribute to approximately 50% of the overall treatment effect. In contrast, we observe that topics related to basic and simple skincare routines lacking high-end or specialized features, such as Topic 10, Topic 18, and Topic 19, do not exhibit any significant increase in their treatment effects. Overall, our findings suggest that experiential store provides benefits to customers, particularly in the context of premium and advanced products. This in turn, positively impacts their purchase decisions. These findings are also in line with those in past work on the impact of customer education initiatives on customer behavior (Bell et al. 2017).

Having estimated the topic-level treatment effects, we now focus on evaluating the relative contribution of specific products to the treatment effect associated with a given topic, which is defined as follows: $w_{ks} = \frac{\hat{\phi}_{ks} Y_s}{\sum_{s=1}^{S} \hat{\phi}_{ks} Y_s}$, where $Y_s = \sum_i \sum_t Y_{it,s}$ denotes the total purchase amount for product s. Figure 7 shows the distributions of the top 10 products that contribute to the purchase amounts of the topics demonstrating the most significant treatment effects. Upon examining the distribution patterns of products within each topic, it becomes evident that Topic 2 exhibits a pronounced concentration of revenue in a limited number of key products. Specifically, the top 10 products within Topic 2 collectively contribute to 74% of the purchase amounts, with the top 5 products playing a substantial role by accounting for over 60% of the total revenue. This concentration underscores the focused and substantial contribution of specific products within Topic 2, highlighting their significant role to overall revenue. Similarly, the purchase amounts in Topic 1 demonstrates a skewed distribution over products, where the top 10 products contribute to about 60% of the total purchase amounts.

Finally, we contrast the results from the aforementioned needs-based analysis with those obtained when products are predefined.[9] By employing the predetermined categorization, we estimate the treatment effects at the category level. The results indicate that experiential store has the most substantial impact on product in the skincare category ($\tau_k = 2.83, p < 0.01$), accounting for approximately 50% of the revenue. This observation is in line with

[9] Following discussions with the partner firm, we decided to use five distinct product categories in our empirical analysis to correspond with the key metrics monitored by the firm regarding customer purchases: skincare, makeup, hair care, bath and body care, and other purchases, wherein fragrances and other categories (e.g., tools, brushes, accessories) were aggregated and grouped.

the results from the needs-based analysis, yet the latter offers more detailed insights into the specific products associated with those particular needs (e.g., Topics 1, 2, 7 and 11) and their relative contribution to the overall revenue. Additionally, experiential store exhibits a significant impact on the makeup category ($\tau_k = 0.89, p < 0.01$), contributing to around 15% of the revenue, while their effects on other categories remain limited.

CONCLUSION

In the ever-evolving retail landscape, businesses are increasingly transforming physical retail spaces into immersive experiential destinations that go beyond the traditional focus on selling products. Naturally, there are questions regarding the economic value of such experiential stores. In collaboration with a company that introduced an experiential store, we use quasi-experimental data over a period of 24 months and measure the causal effect of visiting an experiential store on customer behavior. Our identification strategy involves first creating matched pairs of treated and control groups and then estimating the effects on the sample of matched pairs with the DD approach. We also obtain individual treatment effects by applying GRF with the matched sample. We employ a needs-based modeling framework as well to explore the underlying drivers of our observed effects. Crucially, we connect the two analyses i.e., causal analysis and the analysis of needs, and decompose the overall treatment effect across the differing needs.

We find that experiential store is effective in lifting customer purchases and does so by making treated customers purchase more often and more items. The effect is economically significant and heterogeneous across customers. Interestingly, we observe that the treatment effects largely arise from a small group of customers (20%) experiencing significant positive changes, while a considerable portion of customers shows no significant change. The experiential store has a larger impact on high-value customers (based on past purchase patterns) who actively engaged with the firm but had a longer lapse since their last interaction. Our findings are robust when considering potential confounding effects associated with different time periods, different modeling assumptions, and different treated groups. Finally, using transaction data at both individual and product level, we identify customer needs underlying the effects, their relative contributions to the treatment effect, and the products associated with these needs. Our results show that customer needs exhibiting a significant positive treatment effect are associated with sophisticated skincare routines. These advanced needs are related to high-priced premium products that customers can better assess through hands-on testing and participation in workshops, which are exactly the kinds of activities offered by the store. In contrast, treatment effects associated with needs related to basic and simple skincare routines are not significant. These results are consistent with the experiential learning theory and underscore the significance of haptic experience in acquiring information about products.

Our study presents several important implications for marketing practice. First, our results indicate that offering hands-on experiences and personalized interactions in a retail environment can significantly influence customer preferences. This is particularly true for premium products and specialized needs, such as advanced skincare solutions. However, for product categories that do not rely on tactile experiences, like basic commodities and cleaning supplies, an experiential store may not be a profitable investment. In contrast, categories like luxury watches, cosmetics, and gourmet foods could see significant benefits. This observation is consistent with past work on the demand for products based on search and experience features (e.g., Nelson 1974). Second, our study reveals substantial heterogeneity in the treatment effects among customers, with only a small group of customers showcasing positive effects. This mirrors past findings on the variation in customer lifetime value (e.g., Fader et al. 2022). Our results suggest that firms considering experiential stores should also focus on customer segmentation and targeting. Without this, the return on investment could be limited, as many customers may visit these stores without any significant changes in their purchase patterns. Third, our research relates to the trend of companies engaging with customers early in the buying process, as seen in businesses like Zillow, which aids in home searching and then buying or selling them. Early engagement can be beneficial in educating customers and guiding their purchasing decisions. However, the financial returns from these early-stage touchpoints are often less tangible. Our analysis broadly offers an example of how to assess the value of customer education programs and their impact on the buying process.

As our research is one of the early attempts to identify the causal impact of experiential stores on customer behavior, there is a large potential for future research. First, as our study focused on the effect of an experiential store for a given firm, it is likely that some of our findings could reflect the characteristics of the customer base and product categories of our partner firm. However, it is worth noting that our partner firm is a significant entity within the industry, offering an extensive product portfolio with over $10,000$ products. We hope our research provides a framework for further studies on retail innovations in other product categories. Second, while we document the impact of customer visits to an experiential store on their behavior, we did not have data regarding the specific in-store activities that customers engaged in. Gathering such information could be valuable for companies who wish to decompose the treatment effect across different activities. Related work on shoppertainment has documented the effectiveness of different aspects of livestream shopping, an approach where hosts promote products through live video, albeit within a digital environment (e.g., Liu 2022). Third, our analysis mainly focuses on customer purchases. While sales are an important metric for evaluating the performance of an experiential store, a visit to the store may motivate customers to generate word-of-mouth on digital platforms, which in turn may influence their social connections to visit the store. It will be managerially relevant to segment customers based on their stand-alone value and social value, which captures their impact on peers. Fourth, we use the Latent Dirichlet Allocation model for capturing latent customer needs from customer transactions. There are alternative models like the Hierarchical Dirichlet Process mixture and the mixed membership stochastic block that can be employed as well. Calibrating these latter models in our context, however, is challenging with the large number of products. Finally, it is possible that more companies will begin to introduce physical spaces for enhancing customer engagement. With competition in play, we hypothesize that experience stores that educate customers about firm-specific products may become even more important for garnering their trust. We hope that our work will inspire further studies to deepen the understanding of this nascent and important area of research.

REFERENCES

Angrist JD, Pischke JS (2008) *Mostly Harmless Econometrics: An Empiricist's Companion,* (Princeton University Press).

Ascarza E (2018) Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research* 55(1):80–98.

Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *Annals of Statistics* 47(2):1148–1178.

Austin PC (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine* 28(25):3083–3107.

Bell DR, Gallino S, Moreno A (2020) Customer supercharging in experience-centric channels. *Management Science* 66(9):4096–4107.

Bell SJ, Auh S, Eisingerich AB (2017) Unraveling the customer education paradox: When, and how, should firms educate their customers? *Journal of Service Research* 20(3):306–321.

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *Journal of Machine Learning Research* 3(Jan):993–1022.

Bloch PH, Sherrell DL, Ridgway NM (1986) Consumer search: An extended framework. *Journal of Consumer Research* 13(1):119–126.

Breugelmans E, Altenburg L, Lehmkuhle F, Krafft LL Manfred, Roggeveen AL (2023) The future of physical stores: Creating reasons for customers to visit. *Journal of Retailing* 99(4):532–546.

Cui TH, Ghose A, Halaburda H, Iyengar R, Pauwels K, Sriram S, Tucker C, Venkataraman S (2021) Informational challenges in omnichannel marketing: Remedies and future research. *Journal of Marketing* 85(1):103–120.

Dzyabura D, Jagabathula S (2018) Offline assortment optimization in the presence of an online channel. *Management Science* 64(6):2767–2786.

Fader P, Hardie B, Ross M (2022) *The Customer-Base Audit: The First Step on the Journey to Customer Centricity* (The Wharton School Press).

Forbes (2021) C-suite, it's time to make customer experience and engagement a top priority https://www.forbes.com/sites/forbesbusinesscouncil/2021/05/21/c-suite-its-time-to-make-customer-experience-and-engagement-a-top-priority/?sh=4e9decf87586.

Gauri DK, Jindal RP, Ratchford B, Fox E, Bhatnagar A, Pandey A, Navallo JR, Fogarty J, Carr S, Howerton E (2021) Evolution of retail formats: Past, present, and future. *Journal of Retailing* 97(1):42–61.

Gensler S, Neslin SA, Verhoef PC (2017) The showrooming phenomenon: It's more than just about price. *Journal of Interactive Marketing* 38(1):29–43.

Goldfarb A, Tucker C, Wang Y (2022) Conducting research in marketing with quasi-experiments. *Journal of Marketing* 86(3):1–20.

Goli A, Chintagunta P, Sriram S (2021) Effect of payment on user engagement in online courses. *Journal of Marketing Research* 59(1):11–34.

Grewal D, Levy M, Kumar V (2009) Customer experience management in retailing: An organizing framework. *Journal of Retailing* 85(1):1–14.

Imbens GW, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences,* (Cambridge University Press).

Iyengar R, Park YH, Yu Q (2022) The impact of subscription programs on customer purchases. *Journal of Marketing Research* 59(6):1101–1119.

Jacobs BJD, Donkers B, Fok D (2016) Model-based purchase predictions for large assortments. *Marketing Science* 35(3):389–404.

Jahn S, Nierobisch T, Toporowski W, Dannewald T (2018) Selling the extraordinary in experiential retail stores. *Journal of the Association for Consumer Research* 3(3):412–424.

Kahn BE, Inman JJ, Verhoef PC (2018) Introduction to special issue: Consumer response to the evolving retailing landscape. *Journal of the Association for Consumer Research* 3(3):255–259.

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26(4):379–390.

Kannan PK, Li HA (2017) Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing* 34(1):22–45.

Keele L (2015) The statistics of causal inference: A view from political methodology. *Political Analysis* 23(3):313–335.

Kim M, Zhang J (2023) Discovering online shopping preference structures in large and frequently changing store assortments. *Journal of Marketing Research* 60(4):665–686.

Kolb AY, Kolb DA (2005) Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and Education* 4(2):193–212.

Krishna A (2012) An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology* 22(3):332–351.

Kumar V, Shah D (2004) Building and sustaining profitable customer loyalty for the 21st century. *Journal of Retailing* 80(4):317–329.

Lee L, Ariely D (2006) Shopping goals, goal concreteness, and conditional promotions. *Journal of Consumer Research* 33(1):60–70.

Lemon KN, Verhoef PC (2016) Understanding customer experience throughout the customer journey. *Journal of Marketing* 80(6):69–96.

Levitt T (1980) Marketing success through differentiation—of anything. *Harvard Business Review* 58(1):83–91.

Lin S, Zhang J, Hauser JR (2015) Learning from experience, simply. *Marketing Science* 34(1):1–19.

Liu X (2022) Dynamic coupon targeting using batch deep reinforcement learning: An application to livestream shopping. *Marketing Science* 42(4):637–658.

Manchanda P, Ansari A, Gupta S (1999) The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing science* 18(2):95–114.

Nelson P (1974) Advertising as information. *Journal of Political Economy* 82(4):729–754.

Peck J, Childers TL (2003) To have and to hold: The influence of haptic information on product judgments. *Journal of Marketing* 67(2):35–48.

Robertson T, Gatignon H, Cesareo L (2018) Pop-ups, ephemerality, and consumer experience: The centrality of buzz. *Journal of the Association for Consumer Research* 3(3):425–439.

Rogers EM (2003) *Diffusion of Innovations,* (Simon and Schuster).

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Science* 15(4):321–340.

Seetharaman P, Chib S, Ainslie A, Boatwright P, Chan T, Gupta S, Mehta N, Rao V, Strijnev A (2005) Models of multi-category choice behavior. *Marketing letters* 16:239–254.

Stuart EA (2010) Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1):1–21.

Unal M, Park YH (2023) Fewer clicks, more purchases. *Management Science* 69(12):7317–7334.

Urban GL, Hauser JR (2004) "listening in" to find and explore new combinations of customer needs. *Journal of Marketing* 68(2):72–87.

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113(523):1228–1242.

Wright A, Lynch Jr JG (1995) Communication effects of advertising versus direct experience when both search and experience attributes are present. *Journal of Consumer Research* 21(4):708–718.

Zhang JZ, Chang CW, Neslin SA (2022) How physical stores enhance customer value: The importance of product inspection depth. *Journal of Marketing* 82(2):166–185.

Appendix

These materials have been supplied by the authors to aid in the understanding of their paper.

APPENDIX A: ROBUSTNESS CHECKS

Appendix A provides the robustness of our main findings. First, we investigate the sensitivity of our results concerning the selection of time periods. Second, we explore the robustness of our results related to the specific modeling assumptions. Third, we analyze whether the effects remain similar across different treated groups. Finally, we reiterate the validity of our results through nonparametric estimation. Following the main analysis, we exclude outcome measures from month 0, and robust standard errors, presented in parentheses, are clustered in the analysis.

Alternate Time Periods

A causal effect that is evident in one period may not necessarily apply in another, as the observed effect could be driven by short-term fluctuations or systematic patterns. Additionally, conducting robustness checks with alternate time periods is instrumental in mitigating potential issues related to endogeneity or omitted variable bias. By examining the treatment effect under different timeframes, we can identify and account for confounding factors that might have been omitted from the main analysis.

To investigate whether the treatment effects remain stable or vary across different time periods, we analyze different time specifications, in particular, using 12-month and 18-month data periods in addition to the original 24-month data period. The results, as shown in Table A1, are similar to our main findings. This suggests that the treatment effects remain robust and are not based on the particular selection of time specifications.

Table A1: Treatment Effects by Time Period

	12 months	9 months	6 months
Purchase Amount ($)	5.76***	5.37***	4.70***
	(1.77)	(1.88)	(1.60)
Purchase Frequency	0.10***	0.08***	0.13***
	(0.03)	(0.03)	(0.03)
Purchase Quantity	0.78***	1.02***	0.92***
	(0.27)	(0.34)	(0.30)
Individual fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes
No. of customers	2,098	2,098	2,098
No. of observations	50,352	37,764	25,176

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Alternate Outcomes

When analyzing causal relationships, researchers often make certain assumptions about the distribution of outcomes, and these assumptions can impact the estimated treatment effects. By examining how treatment effects hold up under different outcome transformations, we can assess the robustness of our findings and determine if the observed effects are influenced by the specific modeling choices.

To evaluate the robustness of our findings with respect to different modeling assumptions, as another robustness check, we analyze log-transformed outcomes. Table A2 presents the results, which are qualitatively similar to our primary findings. This suggests that the treatment effects are not based on the specific choice of modeling assumptions.

Table A2: Treatment Effects Using Log-Transformed Outcomes

	Purchase Amount ($)	Purchase Frequency	Purchase Quantity
Mean	0.14***	0.04***	0.07***
	(0.03)	(0.01)	(0.02)
Individual fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes
No. of Customers	2,098	2,098	2,098
No. of Observations	50,352	50,352	50,352

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Alternate Treated Groups

In our main analysis, we focus on a single cohort of the customers who visited the store in January 2020. However, the effects might be based this specific treated group and may not generalize across different subsets of the customers at the firm.

As part of our robustness checks related to selection bias, we extend our analysis to include customers who visited the store during other months, which is presented in Table A3. The last column of Table A3 provides the average treatment effects across all cohorts. Our results indicate that the impact of in-store experience on several outcome measures across multiple cohorts are largely similar, demonstrating the robustness of our findings.

Table A3: Treatment Effects by Cohort

	Dec. 2019	Jan. 2020	Feb. 2020	Mar. 2020	Average
Purchase Amount ($)	4.37***	5.76***	3.86**	3.10*	4.57***
	(1.53)	(1.77)	(1.81)	(1.84)	(0.87)
Purchase Frequency	0.09***	0.10***	0.01	0.05**	0.07***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)
Purchase Quantity	0.55**	0.78***	0.14	0.56**	0.51***
	(0.22)	(0.27)	(0.28)	(0.29)	(0.17)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes
No. Customers	2,864	2,098	1,562	1,166	7,310
No. Observations	68,736	50,352	37,488	27,984	175,440

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Alternate Functional Form

We conduct an additional robustness test to assess the linear and additive structure of our primary estimation. The estimation via GRF allows us to relax this assumption due to its nonparametric nature. As reported in Table 4, our findings are robust.

APPENDIX B: DECOMPOSITION OF THE TREATMENT EFFECT

Appendix B demonstrates that our proposed need-based analysis offers a decomposition of the treatment effect τ into treatment effects across differing needs. To show $\tau = \sum_{k=1}^{K} \tau_k$, note that if $B_{it} = \varnothing$, then $\sum_{k=1}^{K} \mathrm{Y}_{it,k} = \mathrm{Y}_{it} = 0$. Otherwise,

$$\begin{aligned}\sum_{k=1}^{K} \mathrm{Y}_{it,k} &= \sum_{k=1}^{K} \sum_{s \in B_{it}} p(k|s, B_{it}) \cdot \mathrm{Y}_{it,s} \\ &= \sum_{s \in B_{it}} \mathrm{Y}_{it,s} \sum_{k=1}^{K} p(k|s, B_{it}) \\ &= \sum_{s \in B_{it}} \mathrm{Y}_{it,s} = \mathrm{Y}_{it}.\end{aligned}$$

Summing over Equation (4), we have

$$\mathrm{Y}_{it} = \sum_{k=1}^{K} \mathrm{Y}_{it,k} = \left(\sum_{k=1}^{K} \tau_k\right) \cdot \mathrm{W}_{it} + \sum_{k=1}^{K} \theta_{i,k} + \sum_{k=1}^{K} \lambda_{t,k} + \sum_{k=1}^{K} \epsilon_{it,k}. \tag{B1}$$

Given that $\epsilon_{it,k}$ is assumed to be normally distributed with a mean of zero for all $k = 1, .., K$, $\tilde{\epsilon}_{it} = \sum_{k=1}^{K} \epsilon_{it,k}$ is also normally distributed with a mean of zero. As the DD model in Equation (1) is identified, it implies that the DD model in Equation (B1) is also identified. Furthermore, the treatment effects across all $k = 1, .., K$ sum up to the treatment effect such that $\tau = \sum_{k=1}^{K} \tau_k$.

APPENDIX C: LATENT DIRICHLET ALLOCATION

Appendix C provides details on the Latent Dirichlet Allocation (LDA) model which we utilize to infer latent needs. Additionally, we outline the calibration methodology for inferring the optimal number of needs and computing the probability of observing needs based on products and baskets.

Model Specification

Let $\mathcal{S}$ represent the set of non-empty baskets in data. Consider V as the total number of unique products and n_{it} as the number of products in basket B_{it}.

We assume the existence of K needs or topics. Each topic, indexed by $k = 1, \ldots, K$, is associated with a V-dimensional probability vector $\boldsymbol{\phi}_k$ representing the likelihood of each product being associated with topic k. Additionally, each basket B_{it} is linked to a K-dimensional probability vector $\boldsymbol{\theta}_{B_{it}}$ representing the distribution of topics being associated with basket B_{it}.

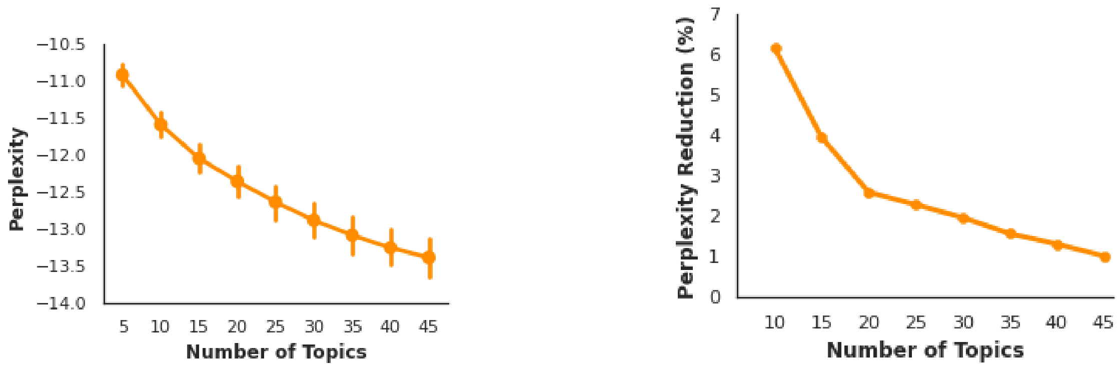
Finally, we assume symmetrical Dirichlet priors for the probability distributions: $\boldsymbol{\phi}_k \sim \text{Dirichlet}(\boldsymbol{\eta})$ for the topics' probability weights over products, and $\boldsymbol{\theta}_{B_{it}} \sim \text{Dirichlet}(\boldsymbol{\alpha})$ for baskets' probability weights over topics. $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ are V- and K-dimensional hyper-parameters, respectively.

LDA assumes that the observed products in our empirical setting stem from a probabilistic generative process as follows:

1. For each topic $k = 1, \ldots, K$,
   (a) Draw a distribution over products $\boldsymbol{\phi}_k \sim \text{Dirichlet}(\boldsymbol{\eta})$.
2. For each non-empty basket $B_{it} \in \mathcal{S}$,
   (a) Draw a distribution over topics $\boldsymbol{\theta}_{B_{it}} \sim \text{Dirichlet}(\boldsymbol{\alpha})$.
   (b) For each product purchase instance $\ell = 1, \ldots, n_{B_{it}}$,
      i. Draw a topic $k_\ell \sim \text{Multinomial}(\boldsymbol{\theta}_{B_{it}})$.
      ii. Draw product $s_\ell \sim \text{Multinomial}(\boldsymbol{\phi}_{k_\ell})$.

Model Estimation and Calibration

We utilize the Gensim package in Python to estimate the model. We set the values of the symmetrical Dirichlet priors parameters to $\boldsymbol{\alpha} = (0.1, \ldots, 0.1)$ and $\boldsymbol{\eta} = (0.1, \ldots, 0.1)$. To determine the optimal number of latent needs, we validate the model using different numbers of topics, ranging from $K = 5$ to $K = 45$, with an increment of five topics at each step. This calibration process involves a five-fold cross-validation. Within each fold, we calibrate the model on 80% of the baskets and assess model perplexity (Blei et al. 2003) using a random 20% holdout split.



(a) Preplexity

(b) Improvement in Perplexity

Figure C.1: Model Performance

Figure C.1a illustrates the resulting perplexity corresponding to each number of latent needs. Lower perplexity values indicate better performance, and we observe an improvement in perplexity as the number of latent topics increases. Additionally, Figure C.1b presents the relative reduction in perplexity. Notably, there is a 6% improvement in perplexity when the number of topics increases from 5 to 10, 4% from 10 to 15, and approximately 2.6% from 15 to 20 topics. However, beyond twenty topics, the reduction in perplexity stabilizes at values below 2%. We therefore retain the 20-topic solution and proceed to retrain the model using the entire corpus of non-empty baskets in our data.

Topic Probability Conditional on Basket and Product

We derive the conditional probability of observing topic k given basket B_{it} and product s. Using Bayes' rule, the probability of observing topic k given product s and basket B_{it} can be expressed as:

$$p(k|s, B_{it}) \propto p(s|k, B_{it})p(k|B_{it}).$$

As per the described data generating process, product s is independent of basket B_{it} conditioned on topic k. Hence, $p(k|s, B_{it})$ simplifies to:

$$p(k|s, B_{it}) \propto p(s|k)p(k|B_{it}).$$

Within the context of the LDA model, $p(s|k)$ is represented by ϕ_{ks} and $p(k|B_{it})$ is denoted by $\theta_{B_{it}k}$. Consequently, $p(k|s, B_{it})$ is given by:

$$p(k|s, B_{it}) \propto \phi_{ks}\theta_{B_{it},k}.$$

In our empirical setting, subsequent to fitting the LDA model, we leverage the estimated values $\hat{\phi}_{ks}$ and $\hat{\theta}_{B_{it},k}$ obtained from the gensim package to compute the conditional topic probabilities.