

The impact of order fulfillment on consumer experience: text mining consumer reviews from Amazon US

Yulia Vakulenko, Diogo Figueirinhas, Daniel Hellström and
Henrik Pålsson

Packaging Logistics, Lund University, Lund, Sweden

Abstract

Purpose – This research analyzes online consumer reviews and ratings to assess e-retail order fulfillment performance. The study aims to (1) identify consumer journey touchpoints in the order fulfillment process and (2) determine their relative importance for the consumer experience.

Design/methodology/approach – Text mining and analytics were employed to examine over 100 m online purchase orders, along with associated consumer reviews and ratings from Amazon US. Using natural language processing techniques, the corpus of reviews was structured to pinpoint touchpoints related to order fulfillment. Reviews were then classified according to their stance (either positive or negative) toward these touchpoints. Finally, the classes were correlated with consumer rating, measured by the number of stars, to determine the relative importance of each touchpoint.

Findings – The study reveals 12 touchpoints within the order fulfillment process, which are split into three groups: delivery, packaging and returns. These touchpoints significantly influence star ratings: positive experiences elevate them, while negative ones reduce them. The findings provide a quantifiable measure of these effects, articulated in terms of star ratings, which directly reflect the influence of experiences on consumer evaluations.

Research limitations/implications – The dataset utilized in this study is from the US market, which limits the generalizability of the findings to other markets. Moreover, the novel methodology used to map and quantify customer journey touchpoints requires further refinement.

Practical implications – In e-retail and logistics, comprehending touchpoints in the order fulfillment process is pivotal. This understanding helps improve consumer interactions and enhance satisfaction. Such insights not only drive higher conversion rates but also guide informed managerial decisions, particularly in service development.

Originality/value – Drawing upon consumer-generated data, this research identifies a cohesive set of touchpoints within the order fulfillment process and quantitatively evaluates their influence on consumer experience using star ratings as a metric.

Keywords Customer journey, Consumer experience, Order fulfillment, Product reviews, Star rating, Delivery, Packaging, Returns

Paper type Research paper

1. Introduction

Providing a superior consumer experience is a key factor in retail success (Lemon and Verhoef, 2016). As a result, the growth of e-commerce has transformed the requirements placed on logistics and operational capabilities. Nowadays, order fulfillment services



increasingly focus on meeting consumers' expectations in the transition toward consumer-centric supply chain management (Esper *et al.*, 2021). This is because the performance of the order fulfillment service mediates the consumers' satisfaction with the whole retail experience, translating into consumer loyalty. Lambert (1992, p. 18) indicated that "the starting point in developing a logistics strategy must be a thorough understanding of final customers' requirements. Only then is it possible to determine the required performance of firms throughout the supply chain." To gain competitive advantage, there are opportunities for actors in e-commerce to incorporate the consumer perspective in the evaluation and design of order fulfillment services. This is increasingly possible as consumers generate vast amounts of digital footprints during their online shopping and post-purchase experiences. In e-retail, the multi-stage approach to consumer experience co-created by multiple actors is referred to as the e-customer journey (Lemon and Verhoef, 2016; Vakulenko *et al.*, 2019).

An e-customer journey constitutes all the steps, decision-making occasions and various types of responses a consumer goes through when making a purchase, including interactions with the e-retailer and fulfillment providers. These customer journey nodes are referred to as touchpoints, which can be represented by service variables like delivery price, delivery time, delivery points' locations, delivery accuracy, return policies and delivery speed (Mentzer *et al.*, 2001; Rao *et al.*, 2011a). From the moment of order placement, consumers follow a chain of service events and decision-making nodes (i.e. touchpoints) that directly affect their (dis)satisfaction and relationship with the retailer and fulfillment service provider. Traditionally, it has been challenging for e-retailers to gain comprehensive insights into consumers' delivery experiences, as fulfillment and logistics services focused on operational service quality measures. From the academic perspective, previous studies in consumer research in delivery experience have been largely based on surveys. Their common drawbacks include a lack of insights into some consumer experience dimensions, limited capacity to reflect the continuous experience perspective (De Leeuw *et al.*, 2012), potential data quality issues (Moy and Murphy, 2016) and reliance on human informants (Speklé and Widener, 2017). The customer journey approach is a comprehensive take on mapping and understanding consumers' order fulfillment experiences, which has the potential to identify and determine critical touchpoints for order fulfillment services, specifically in combination with operational empirical data that can capture a continuous nature of the consumer experience.

Using consumer data and feedback can greatly improve forecasting, inventory, delivery and service design to meet consumer needs more effectively and enhance overall efficiency. The growing importance of consumer experience is pushing managers and academics to utilize new data sources and analytics to gain valuable insights. The spread and rapid availability of online services and the ubiquity of technology have brought a stream of user-generated data whose magnitude has increased immeasurably. A visible part of such footprint exists in the form of electronic word-of-mouth (eWOM), such as star ratings and consumer reviews, which are used by both retailers (Floyd *et al.*, 2014) and consumers (Mudambi and Schuff, 2010; Ngo-Ye and Sinha, 2014). Star ratings and reviews influence purchasing decisions, as research shows their direct impact on sales (Floyd *et al.*, 2014; Wang *et al.*, 2015) and product returns (Minnema *et al.*, 2016), positioning them as performance indicators and essential management tools. For consumers, online reviews have become an indispensable tool to navigate in a retail space flooded with an overwhelming number of choices. Depending on the region, 33%–75% of online shoppers find reviews helpful (Statista, 2024) and over 90% read reviews before making a purchase decision (Statista, 2021). Moreover, products with an average star rating of four stars and above account for over 45% of e-commerce site page traffic worldwide (Statista, 2022).

This paper adopts a novel approach to understanding consumers' needs and behavioral responses through a theoretical lance of the customer journey in combination with natural language processing techniques for big data analysis. The study examines e-consumer feedback as a form of consumer-generated empirical data to test its suitability for evaluating the order fulfillment performance in e-retail. The purpose and design of this study are informed and motivated by the rise of consumer-centricity in logistics and supply-chain management (Esper *et al.*, 2021), the latest developments in natural language processing techniques and the increasing availability of big datasets in the studied field. The study aims to (1) identify order fulfillment touchpoints in the e-consumer journey and (2) determine their relative importance for the consumer experience. To do that, the study adopts a big dataset from a major online retail platform and text mining techniques.

2. Theoretical background

2.1 Order fulfillment process

Order fulfillment is an essential process in logistics and supply chain management and a critical aspect of customer service. This process comprises generating, filling, processing, delivering and servicing customer orders (Croxtton, 2003; Waller *et al.*, 1995). In other words, it starts with the consumer placing an order and ends with the consumer (or customer) receiving the ordered products or services. Shapiro *et al.* (1993) argue that through this process, the consumer interacts with the firm, and therefore order fulfillment determines their experience with the firm. From a management point of view, the order fulfillment process overlaps several functional responsibilities and covers a series of operations (i.e. inventory management, inventory storage, receiving orders, picking, packaging, shipping and returns management). These crucial operations can turn order fulfillment into a competitive advantage as an effective process ensures that consumers receive their orders accurately (e.g. in the right quantity and in the right condition) and in a timely manner (Christopher, 2018; Taylor *et al.*, 2019).

In e-retail, order fulfillment is an integrated part of the e-fulfillment process. This process includes overall business strategy and website performance (Bagdare and Jain, 2013; Titiyal *et al.*, 2019). An e-fulfillment is commonly characterized by business quality (strategy, product, transaction and website quality), distribution (assortment width and type and delivery time and area), last mile delivery (flexibility, reliability, information, time and cost) and returns management (recovery strategy, return policy and consumers' effort). This makes the operational dimensions of e-fulfillment a focal point for retailers and logistics service providers. Notably, as a part of e-fulfillment, order fulfillment offers managers a great opportunity to improve overall operations and the consumer experience.

Effective fulfillment design and operations depend on integrating consumer insights to optimize demand forecasting, inventory management, warehouse management and delivery strategies, ensuring alignment with consumer needs. In demand forecasting, consumer insights enable more accurate prediction in demand, thus reducing the risks of over- and understocking and facilitating better inventory management (Schaer *et al.*, 2019). Furthermore, in inventory management, consumer feedback provides insights to enable strategic inventory placement (i.e. to be closer to the consumer and prioritize inventory location based on service cost and delivery distance), which is crucial in high demand seasons. In warehouse management, consumer insights can support efficient algorithm design for order packing and picking and in transport and delivery, consumer insights allow strategic decision-making on service feature prioritization, such as cost, speed, location, delivery mode and degree of the process transparency (Vakulenko *et al.*, 2019). Finally, in returns management, consumer experience insights can help establish a return policy and

strategy that is more operationally and cost-efficient (Minnema *et al.*, 2016). With the rising need for understanding and managing consumer experience, drawn by its impact and potential toward operations in order fulfillment and supply chain management in general, managers and academics are turning toward new sources of data and data analytics to extract knowledge from the petabytes of information consumers leave online (Erevelles *et al.*, 2016).

2.2 Customer journey touchpoints

A consumer-centric approach to operations and supply chain management is grounded in the notion that operational capabilities and performance need to correspond to consumer preferences, needs and experiences (Esper *et al.*, 2021). It means that decisions and performances in a supply chain, particularly in the order fulfillment process, affect consumer experience, which directly impacts customer satisfaction and the brand image and consequently, carries significant implications for businesses. To enhance the consumer experience, businesses need to fully grasp the journey their customers go through. The concept of the customer journey implies that the consumer experience is a continuous process connected through past, present and future interactions, where each individual experience can be split into pre-purchase, purchase and post-purchase stages (Lemon and Verhoef, 2016). As consumers go through each stage of their journey, they encounter various interactions, responses and decision-making instances, commonly referred to as touchpoints, which translate into satisfaction or dissatisfaction. Customer journey touchpoints can vary widely from a marketing campaign, website visit, store experience, to a customer support call and are categorized into four groups: brand-owned, partner-owned, customer-owned and the external sources (i.e. experience environment and social nodes).

Retail environments present numerous critical touchpoints, each playing a role in shaping experiences that add up to how consumers make decisions and shape their behavior (Baxendale *et al.*, 2015; Verhoef *et al.*, 2009). The e-commerce customer journey within the control of retailers and logistics service providers can be distilled into clusters of touchpoints, including marketing communication, web-shopping experiences, order delivery and returns. Each cluster comprises various touchpoint categories that differ among retailers, the products they offer, payment options, delivery services and other experiential aspects. Fulfillment-related touchpoints have been studied to varying degrees and frequencies, largely influenced by the advancements of the Logistics Service Quality (LSQ) and Physical Distribution Service Quality (PDSQ) frameworks (Mentzer *et al.*, 1989, 2001; Rao *et al.*, 2014). Traditional dimensions of these experiences include delivery price (Rao *et al.*, 2011a), delivery options, delivery speed (Andrejić, 2019) and the range of delivery failures (Rao *et al.*, 2011b). Another significant set of touchpoints concerns the return experience, explored through consumer research on return costs and policies (Duong *et al.*, 2022). Notably, a broad body of studies has examined consumer experience using a symptomatic approach, leveraging operational data to understand consumer behavioral response to specific operational touchpoints, such as return volumes, return frequencies, order characteristics and consumer retention (Griffis *et al.*, 2012; Shang *et al.*, 2019). Additionally, other aspects of consumer experience in retail, falling under the responsibility of fulfillment service providers, have not been traditionally included or investigated as fulfillment experience touchpoints. Examples include the brand image and the packaging experience (Joutsela *et al.*, 2016). Specifically, while packaging significantly impacts consumer experience through a range of packaging features and performances (Dash, 2021), its role in shaping the perceived fulfillment experience remains unclear.

In the e-retail landscape, where the brand perception is blended with the retail function, distinctions between brand-owned and partner-owned touchpoints become less clear (Vakulenko *et al.*, 2019). The use of third-party logistics and last-mile delivery partners are instances of partner-owned touchpoints where e-retailers do not maintain full control. Existing literature, such as Yu *et al.* (2015), shows that the choice of a delivery service provider by an e-retailer influences customer satisfaction. However, few studies have systematically identified the associated touchpoints and quantified their impact. Consequently, many e-retailers and logistics service providers may not fully grasp the profound impact that third-party logistics and last-mile delivery performance can have on consumer experience.

2.3 Star ratings and reviews

The nature of rating and reviewing behavior is shaped by various psychological phenomena (e.g. cautious and subconscious triggers that translate into needs, emotions and behavioral impulses), social trends (e.g. consumption trends) and environmental/situational triggers (e.g. service or product features) (Hernández-Ortega, 2018; Hoffart *et al.*, 2019; Koh *et al.*, 2010). These factors are critical to how consumers perceive and respond to service offers and events, how eWOM is perceived and utilized and which aspects are reflected in eWOM like ratings and reviews (Chen and Xie, 2008; Chevalier and Mayzlin, 2006). This results in different rating patterns, including how ratings affect purchases, the timing of consumer ratings, the influence of positive and negative product and service features on ratings, rating distribution and the impact on future purchases (Anderson and Simester, 2014; Hernández-Ortega, 2018; Sunder *et al.*, 2019).

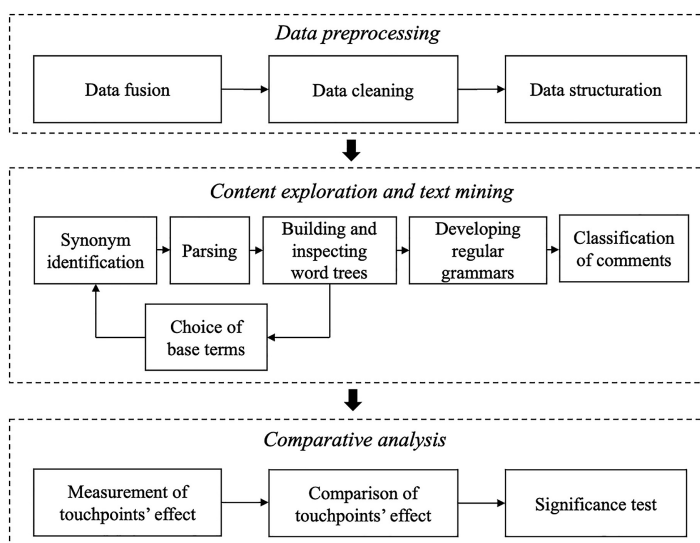
Different stages of consumer interactions with eWOM, in the form of ratings and reviews, can carry different theoretical connotations and explanations. The initial interaction stage, where consumers receive and interpret information that later translates into purchasing behavior and further eWOM communication, stems from the psychological and behavioral processes that can be captured by Spence (1973) in signaling theory. Here, consumers seek to resolve information asymmetry by gaining insights into a business's offers through ratings and reviews. Furthermore, by combining information from existing eWOM or entering the e-retail experience solely with preformed needs, consumers benchmark their experiences against expectations. This benchmarking process leads to different levels of satisfaction and loyalty, explained by the confirmation–disconfirmation paradigm from cognitive dissonance theory (Festinger, 1957) and the expectancy confirmation paradigm (Oliver, 1977). This paradigm predicts customer satisfaction based on expectations and experiences, measuring the difference as disconfirmation or cognitive dissonance. Negativity bias from psychology can also help to explain consumer behavior. Negativity bias refers to the cognitive effect where negative events and experiences have a greater impact on an individual's thoughts, feelings and behavior than positive ones (Rozin and Royzman, 2001). Customers are more likely to remember and be influenced by negative experiences, leading to stronger and longer-lasting impressions. While a positive bias also exists, it is generally weaker and requires more repetition to influence behavior (Cacioppo *et al.*, 1999; Peeters, 1991). Related to negativity bias is the concept of loss aversion, which suggests that individuals usually prefer to avoid losses rather than acquire equivalent gains (Kahneman and Tversky, 1979). In other words, individuals derive more satisfaction from not losing something than from gaining something of similar value. The negativity bias and loss aversion offer possible explanations for rating and reviewing behaviors, where most consumer engagement is linked to noticing and responding to the most positive and most negative product and service features.

User ratings influence purchase intention by affecting the perception of product quality (Flanagin *et al.*, 2014) and the actual purchasing decisions, as research shows their direct

impact on sales (Floyd *et al.*, 2014; Wang *et al.*, 2015) and product returns (Minnema *et al.*, 2016), positioning them as performance indicators and essential management tools. Furthermore, star ratings can be seen as an aggregate reflection of the consumers' product perception and its performance combined with the brand reputation, price and average score of previous ratings (Engler *et al.*, 2015). In practice, review systems result in an overall rating, combining evaluations of products and related services (Bhatt *et al.*, 2015). Integrating service and product experiences in reviews, which can include reflections on e-service quality components, such as website design, security, information quality or fulfillment, shows how star ratings and reviews are a construct of multiple stages and touchpoints of consumer experience. Therefore, star ratings and consumer reviews can be used in combination to develop an in-depth understanding of the consumer's journey.

3. Methodology

We conducted Exploratory Data Analysis (EDA) using text mining techniques to explore the review dataset and identify key touchpoints. EDA is a crucial initial phase in the knowledge discovery process, allowing scientists to explore unfamiliar data through a series of analysis operations, such as filtering, aggregation and visualization (Milo and Somech, 2020), and EDA should be seen as an integral part of statistical inference and model-building (Tukey, 1977). Our methodology followed the standard process for text mining studies (Antons *et al.*, 2020), which includes data gathering, text processing, content analysis and integration with the study. We introduced the use of word trees in the exploratory phase to identify relevant topics and patterns for constructing regular grammars for later classification as a means of adopting the EDA research framework to this study. An overview of this process is illustrated in Figure 1. The research employs text mining techniques to analyze online consumer reviews and ratings to assess e-retail order fulfillment performance. Text mining refers to techniques for transforming unstructured text data into structured data, enabling the application of mathematical and statistical methods (Miner, 2012).



Source(s): Figure created by the authors

Figure 1.
Research process

The dataset utilized for this study was obtained from the Amazon Customer Review Library, a publicly available platform adhering to Amazon's terms of use, from an international online retail platform, Amazon US. The US was identified as a good fit for this investigation due to its well-established and mature e-commerce market, coupled with the rapid expansion of its logistics sector that provides highly competitive services to online retailers. Additionally, the US e-retail market is distinguished by its long-standing practice of incorporating consumer feedback through ratings and reviews. Given that Amazon US is a dominant actor in the US e-commerce sector, with both its own retail platform and logistics operations, the insights from the consumer ratings and reviews can provide a comprehensive examination.

The dataset includes consumer reviews, user ratings and retailer and product information from 1995 to 2015. Despite being nearly a decade old, the dataset used in this study is invaluable due to its unique characteristics, size and quality. Provided by Amazon, it remains an unparalleled source of rich information, notable for being in the public domain. Furthermore, utilizing this unique dataset allowed us to demonstrate the value of the methods employed to extract information volunteered by consumers embedded within the reviews. Consumers are invited to leave feedback upon delivery, allowing them to express their opinions and describe their experiences regarding products purchased from [Amazon.com](https://www.amazon.com). Prior to the application of data mining methods, data preprocessing methods were used to deal with the data imperfections, which often contain inconsistencies, redundancies or superfluous information, rendering the low data quality for knowledge extraction ([García et al., 2016](#)). [García et al. \(2016\)](#) noted that "data preprocessing is able to adapt the data to the requirements posed by each data mining algorithm, enabling to process data that would be unfeasible otherwise."

3.1 Data preprocessing

We applied the classical steps of data preprocessing – data fusion, data cleaning and data structuration ([Tanasa and Trousse, 2004](#)). In the data cleaning step, we removed faulty entries, which included incorrect data types, incorrect numbers of fields or missing values. Datasets in languages other than English were excluded, as were reviews of products delivered digitally. Additionally, to avoid the problem of fraudulent reviews ([Hu et al., 2012](#); [Wu et al., 2020](#)), the study included only reviews from verified purchases, which excluded about 20% of all reviews, resulting in a total of 100.7 m data entries. The number of reviews per user was also controlled to prevent a skewing of the results due to non-representative users. Duplicates were then eliminated, leaving 98.9 m reviews for analysis.

3.2 Content exploration and text mining

Text mining is a technique used to extract information from textual data, such as social network feeds, emails, blogs, online forums, survey responses, corporate documents and news ([Gandomi and Haider, 2015](#)). The information extraction techniques used in text mining and natural language processing enable the retrieval of structured information from unstructured textual data ([Gandomi and Haider, 2015](#)). We devised a method to detect and classify information about order fulfillment that was buried within the product review comments. The sequence of steps followed is shown in [Figure 1](#), including the following steps:

Choosing base terms. An initial list of words referring to the order fulfillment process was selected based on a review of studies and conceptual frameworks related to fulfillment, delivery and return management in B2C e-retail. Guided by the Logistics Service Quality (LSQ) and Physical Distribution Service Quality (PDSQ) frameworks ([Mentzer et al., 1989, 2001](#); [Rao et al., 2014](#)), terms such as availability, timeliness, condition and accuracy were considered. For product returns, terms like time, cost and effort were incorporated based on the work by [Janakiraman et al. \(2016\)](#).

Identifying synonyms. Synonyms for each base term were identified using tools such as WordNet (Fellbaum, 1998; McCrae *et al.*, 2020).

Parsing. The whole review corpus was sentence tokenized, that is, segmented into sentences and words and parsed for sentences containing the keywords.

Building and inspecting word trees. The body of sentences found in the parsing step was subsequently organized into word trees. Word trees are visual text summaries created using syntax trees that aggregate sentences by their shared words and split those sentences into branches at the points where they diverge (Nan and Cui, 2016). The size of the words or word combinations in the word trees is displayed according to frequency, with their sizes being proportional to their frequencies, as is done in word clouds. The interactive nature of the word trees was utilized to navigate through the tree-like structure to identify and focus on the most frequently appearing words and word combinations [1]. This method permitted efficient detection of the types of comments made regarding the order fulfillment touchpoints within the review comments as well as the manner in which they were expressed and their grammatical structure. The expressions were collected and compiled into two categories: positive order experiences and negative order experiences. Additionally, this process facilitated the identification of new base terms, resulting in an iterative discovery process.

Developing regular grammars. The information gathered in the aforementioned step served as the basis for constructing regular grammars and bag-of-words for each part-of-speech. These regular grammars were later compiled as regular expressions, a general pattern notation that allows the encoding of text patterns (Friedl, 2006). Using WordNet (Fellbaum, 1998), synonyms were incorporated into the regular expressions, which, with the use of spelling error corpora, also accounted for typical spelling mistakes. This procedure enabled emulating word stemming without the need to edit the whole review corpus.

The specificity of the expressions of interest in this study made regular expressions a well-fitting tool to encode the patterns identified through the word trees. By using regular expressions, tight control over the classified comments was possible since all expressions are strictly defined, ensuring transparency with a well-defined set of search expressions and avoiding the black-box nature and low accuracy of machine learning tools.

Besides the increased control over the search patterns, by not resorting to tools such as part-of-speech (POS) tagging or machine-learned sentiment analysis, the analysis avoided inaccuracies in word tagging and sentiment classification, which could result in misclassifying comments or failing to classify them. This outcome is particularly true for the texts under consideration, whose conversational nature entails a “lack of conventional orthography, noise, linguistic errors, spelling inconsistencies, informal abbreviations, and idiosyncratic style” (Meftah *et al.*, 2018, p. 2821). A good example of such issues can be found in the work of Gimpel *et al.* (2010), which showed how the accuracy of the Stanford POS tagger trained on the Wall Street Journal drops from 97% accuracy in standard English to 85% in tweets (Toutanova *et al.*, 2003). State-of-the-art learned methods for POS tagging social network texts perform at about 90%, which means one in ten words are misclassified (Goh *et al.*, 2022; Meftah *et al.*, 2018), resulting in sentence accuracy varying from 57% to 23% (Derczynski *et al.*, 2013), with the latest state-of-the-art methods achieving sentence accuracies of around 76% (Li *et al.*, 2021). It is also worth noting that the high reported accuracy of the best models was obtained from evaluations on the Wall Street Journal section of the Penn Treebank, the de facto standard for evaluating POS taggers and perform significantly worse when used in out-of-domain, with accuracies decreasing to around 80% across domains (Hansen and van der Goot, 2023). In a similar way, using sentiment analysis for classifying the valence of comments, which is extremely context sensitive (Rambocas and Pacheco, 2018), would carry accuracy costs when using pre-trained classifiers or entail the same or more manual labor in annotating training data (Pandey *et al.*, 2022) as crafting regular expressions if building custom dictionaries, while still carrying the accuracy costs of

POS tagging mentioned earlier. In the end, at the cost of more laborious work defining the regular grammars and bag-of-words for each part-of-speech involved for each class, the analysis procedure minimized the risk of misclassification (false positives), since regular expressions are strict patterns, providing sampling large enough for high confidence levels with small margins of error.

Classifying comments. Using the regular grammars, each comment underwent a scanning to determine its classification within one of the touchpoint categories as well as its polarity, whether positive or negative.

3.3 Comparative analysis

In this section, we use multivariate hypergeometric analysis to measure and compare the effects of touchpoints on ratings. This analysis helps establish the strength of both negative and positive experiences regarding each touchpoint and their relative importance.

Measurement of touchpoints' effect. After classifying all the reviews, we conducted an analysis to measure and compare the effects of the identified touchpoints on ratings. We began by plotting and examining the rating distributions of both negatively and positively classified reviews for each touchpoint. We then compared these observed distributions to the expected distributions for random samples of the same size. Using multivariate hypergeometric (MVHG) analysis (Bishop *et al.*, 2007), we determined whether the observed values were within expected values and could be attributed to randomness or if the deviation fell outside any reasonable expectation. In the latter case, the difference represents the size of the effect. Products with exclusively positive or negative reviews were excluded, as there would be no contrasting reviews for comparison. Hence, the comparison focused on products of the same kind that received both positive/negative and non-positive/non-negative experiences.

Comparison of touchpoints' effect. Once the effects were measured, we compared them to investigate the relative weight and importance of the identified touchpoints on the star ratings. Additionally, we tested the presence of these effects within each product category.

Significance test. To determine the significance of the observed patterns, the study investigated if the over/underrepresentation in star ratings could be attributed to random variation. The examination of the rating distribution utilized principles of discrete multivariate hypergeometric distributions (MVHGD), as detailed in [Appendix 1](#). Given the extreme memory and computational demands of cumulative multivariate hypergeometric calculations, we restricted the tests to over/underrepresentation of one and five-star ratings. This approach provided a conservative upper bound for the p -values, yet it was still sufficient to validate the effect's existence beyond a reasonable doubt.

4. Results

The study identified 12 touchpoints within the order fulfillment process that impact consumer experience. These touchpoints span three core categories – delivery, packaging and returns (see [Table 1](#)). The star rating distribution for positively and negatively classified reviews for each touchpoint is presented in [Table 2](#). The results show that positive experiences, including fast delivery, free returns and multiple delivery options, translate into positive reviews and higher star ratings. In contrast, negative experiences, such as damaged packages, misdelivered items and late deliveries, result in negative reviews and reduced ratings. The p -values presented in [Table 2](#) show that reviews classified as positive significantly surpass the aggregate ratings across six touchpoints: delivery process, speed, timeliness, options, cost and packaging. This highlights the critical role these touchpoints play in enhancing customer satisfaction. In contrast, for delivery information, delivery

Touchpoint category	Touchpoint	Touchpoint definition	Touchpoint description	Examples from online reviews
Delivery	Delivery process (overall delivery service experience)	Reviews have contained non-detailed descriptions of the delivery service encounter	As a touchpoint, the delivery process refers to the entire sequence of steps involved in delivering the product to fulfill an order. A positive delivery service experience is efficient and can be characterized by timely and speedy delivery, order accuracy, fair price, availability of information about the delivery and suitable delivery options as well as minimized frictions throughout the delivery process	<i>good delivery, terrible delivery, satisfied with the delivery, no problems with the delivery, etc.</i>
	Delivery speed	Consumers' perception of the delivery length	Consumers expect fast shipping options and respond positively to shorter delivery times. The delivery speed touchpoint is the experiential node that refers to the actual speed offered at the moment of order placement, its consequential fulfillment and, most importantly, the consumers' perception of the delivery speed	<i>slow delivery, long shipment, fast delivery service, etc.</i>
	Delivery cost	Consumers' perception of delivery service price	Consumers are likely to consider the cost of shipping when making purchasing decisions and will refer to the value-for-price perception when evaluating the delivery experience and whether it is worth the price placed on the service. Consumer perception of the delivery price (or delivery cost) is a function of the actual service price and the consumer's interpretation of the fairness/suitability of that price	<i>expensive delivery, have to pay for shipment, freed shipment, etc.</i>
	Delivery order accuracy	Completeness and accuracy of the order	Delivery order accuracy refers to the accuracy of the order, including the correctness of the item(s) (i.e. ensuring that the items placed in the order were not replaced with other items during delivery), the completeness of the order (i.e. ensuring that all the items and their parts from the order are not missing) and the correct color and size of the ordered item. Generally, potential issues are linked with the process of order picking. However, some cases are linked to the quality and reliability of the presentation of the item on the website	<i>wrong size, wrong color, missing an item, etc.</i>
	Delivery information	Information availability during the delivery process	Delivery information represents consumers' experience and perception of the availability of necessary information during the delivery process. The information outlet can take the form of accurate order tracking, order placement confirmation, delivery process notifications and delivery confirmation. Notably, an oversupply of delivery information in the form of excessive notifications can translate into a negative experience and consumer dissatisfaction	<i>provided confirmation, wrong tracking number, tracking not working, etc.</i>
	Delivery on time	Fulfillment of the delivery deadline	On time delivery represents consumers' response to the fulfillment of the promised delivery deadline (i.e. whether the order arrived on time or was late based on the promise at the time of order placement)	<i>delivery came late, order came earlier than promised, arrived on time, etc.</i>
	Delivery options	Availability of desirable and/or multiple delivery options	Delivery options, as an experience touchpoint, refer to consumers' (dis)satisfaction with the provided delivery options. This can be based on the number of available delivery options (i.e. whether multiple options were provided and/or whether there were enough delivery options), the availability of fast delivery options such as the same-day delivery or next-day delivery and/or a common standard of what is perceived as fast delivery. Additionally, the availability of cheap or free delivery options, the availability of the desired delivery modes (e.g. home delivery, office delivery, etc.) and the possibility to choose a specific/alternative delivery address	<i>good delivery, terrible delivery, satisfied with the delivery, no problems with the delivery, etc.</i>

Table 1.
Description of order
(continued) fulfillment touchpoints

Touchpoint category	Touchpoint	Touchpoint definition	Touchpoint description	Examples from online reviews
Packaging	Packaging	Consumers' packaging perception and experience	Packaging, as a touchpoint, refers to the consumer's interaction with primary packaging and, in some cases, secondary packaging during the process of order receiving, transportation to the point of consumption and unpacking. This interaction can be related to different packaging features, such as the package being open or intact at the moment of delivery, the quality of the actual packaging, a general broad perception of the packaging and, most frequently, the package's condition upon delivery (i.e. whether it was damaged in any way)	<i>slow delivery, long shipment, fast delivery service, etc.</i>
Return	Return process (overall return service experience)	Non-detailed descriptions of the return service encounter	As a touchpoint, the return process refers to the entire sequence of steps involved in returning the product. A positive return service experience is efficient and can be characterized by minimal effort, clarity of the process, fair price and return policies as well as minimized frictions throughout the return process	<i>expensive delivery, have to pay for shipment, freed shipment, etc.</i>
	Return cost	Consumers' perception of the return price	The return cost touchpoint represents consumers' perception of the return cost/price and the evaluation of the appropriateness of that cost, whether free or falls within different price ranges	<i>wrong size, wrong color, missing an item, etc.</i>
	Return effort	Consumers' perception of the amount of effort required to perform the return	Return effort represents consumers' perception and evaluation of the amount of effort required to return the order and how appropriate this effort was. Naturally, a return experience requiring a lot of effort translates into consumer dissatisfaction. This can be rooted in return instructions that are hard to retrieve, return declarations, finding and printing return labels for shipping, waiting for and receiving reimbursements, taking the order to the dedicated logistics destination/hub and figuring out which packaging can be used for shipment and how to use it	<i>provided confirmation, wrong tracking number, tracking not working, etc.</i>
	Return policy	Consumers' perception of the return policy fairness	As a touchpoint, the return policy represents consumers' perception of its fairness, suitability and clarity. The evaluations and resulting content of the reviews are provided by both returning consumers and consumers who chose not to return the product due to an unsuitable policy. These reviews can pertain to return deadlines, return fees and other conditions associated with the return rules	<i>delivery came late, order came earlier than promised, arrived on time, etc.</i>

Table 1. Source(s): Table created by the authors'

accuracy and all touchpoints concerning returns, there's a noticeable decline in the overall star rating, even for the reviews categorized as positive.

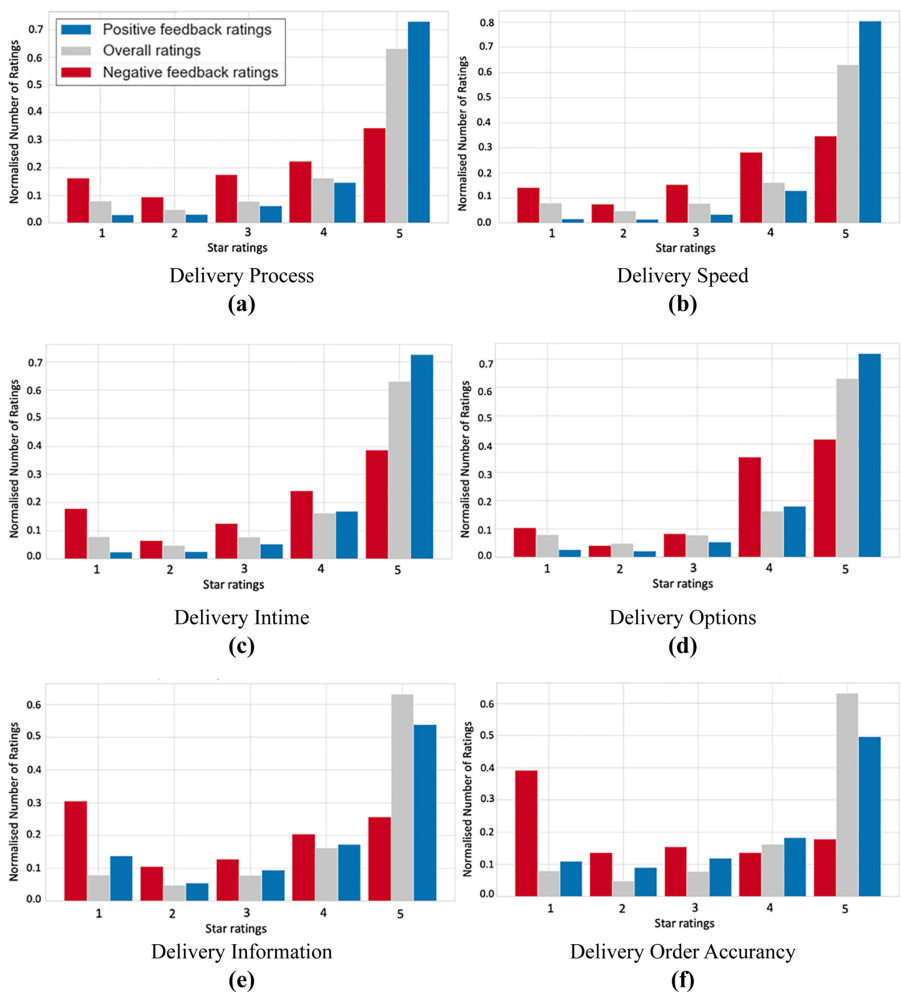
4.1 Distribution of customer ratings

Traditionally, customer ratings tend to follow a “J-shaped” distribution, where the bulk of ratings typically cluster at the five-star level, with fewer ratings at each subsequent lower star level and a slight increase in one-star ratings (Hu *et al.*, 2009). Figure 2 depicts the normalized rating distributions: overall ratings in gray, positive experiences in blue and negative experiences in red. Notably, across all touchpoints, reviews classified as positive consistently outperformed negative ones, with a higher prevalence of five-star ratings and fewer one-star evaluations. Based on the distribution in customer ratings shown in Figure 2, the following sub-sections delve into the distribution of ratings for touchpoints related to delivery,

Touchpoint	Av. star rating pos	Av. star rating neg	Av. Diff pos	Av. Diff neg	p-value pos	p-value neg	Distribution "positive"	Distribution "negative"
Delivery process	451	3.49	0.29	-0.73	0.000	0.000	[3,836, 3,967, 7,868, 18,554, 92,043]	[1,071, 624, 1,151, 1,470, 2,261]
Delivery speed	469	3.62	0.47	-0.6	0.000	0.000	[15,542, 14,927, 33,000, 123,483, 774,152]	[532, 283, 578, 1,064, 1,304]
Delivery on time	455	3.59	0.33	-0.62	0.000	0.0000	[1,182, 1,249, 2,551, 8,115, 34,635]	[494, 180, 347, 669, 1,068]
Delivery options	454	3.94	0.32	-0.28	0.000	0.0019	[49, 41, 99, 334, 1,335]	[5, 2, 4, 17, 20]
Delivery info	392	3	-0.3	-1.22	0.000	0.000	[359, 143, 247, 449, 1,401]	[291, 101, 122, 195, 245]
Delivery order	386	2.57	-0.35	-1.65	0.000	0.000	[818, 674, 890, 1,356, 3,689]	[23,923, 8,332, 9,417, 8,303, 10,875]
accuracy								
Delivery cost	446	2.19	0.24	-2.03	0.000	0.000	[333, 164, 275, 991, 4,431]	[1,111, 282, 216, 202, 298]
Packaging	446	2.86	0.25	-1.36	0.000	0.000	[10,366, 8,248, 14,918, 41,399, 181,090]	[21,449, 10,643, 14,264, 14,889, 13,946]
Return process	325	2.34	-0.97	-1.87	0.000	0.000	[1893, 1,140, 1,026, 1,371, 2,801]	[1,128, 409, 298, 203, 423]
Return cost	307	1.72	-1.15	-2.5	0.000	0.000	[2,509, 1,508, 1,426, 1,609, 2,781]	[7,587, 2,150, 1,253, 667, 535]
Return effort	287	2.41	-1.35	-1.81	0.000	0.000	[2,556, 1,548, 1,375, 1,433, 2025]	[360, 168, 128, 106, 127]
Return policy	268	1.58	-1.54	-2.64	0.000	0.000	[3,628, 1,623, 1,145, 1,084, 2,332]	[535, 116, 57, 31, 31]

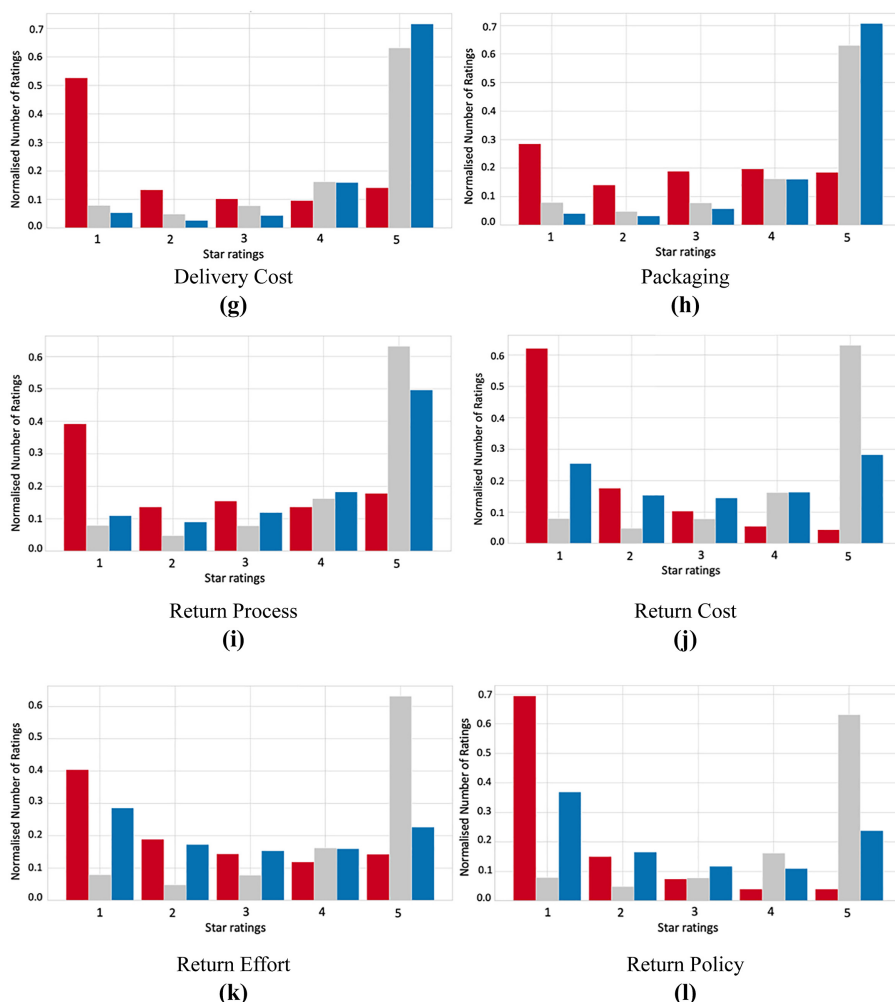
Source(s): Table created by the authors

Table 2.
Order fulfillment
touchpoints' effect on
online user ratings



(continued)

Figure 2.
Rating distributions
across order fulfillment
touchpoints



Source(s): Figure created by the authors

Figure 2.

packaging and returns, shedding light on their impact on both positive and negative consumer experiences. Finally, the sub-section summarizes how the 12 order fulfillment touchpoints influence star ratings.

4.1.1 Influence of delivery touchpoints on ratings. The study reveals seven essential delivery touchpoints, demonstrating a correlation between the nature of delivery experiences and subsequent consumer ratings. These touchpoints include the delivery process/procedure, speed, timeliness, options, information, accuracy and cost. Consistency across all delivery touchpoints is observed, with positive delivery experiences generally correlating with higher star ratings, whereas negative delivery experiences are associated with lower ratings. Notably, the touchpoints of delivery information and order accuracy show lower-than-

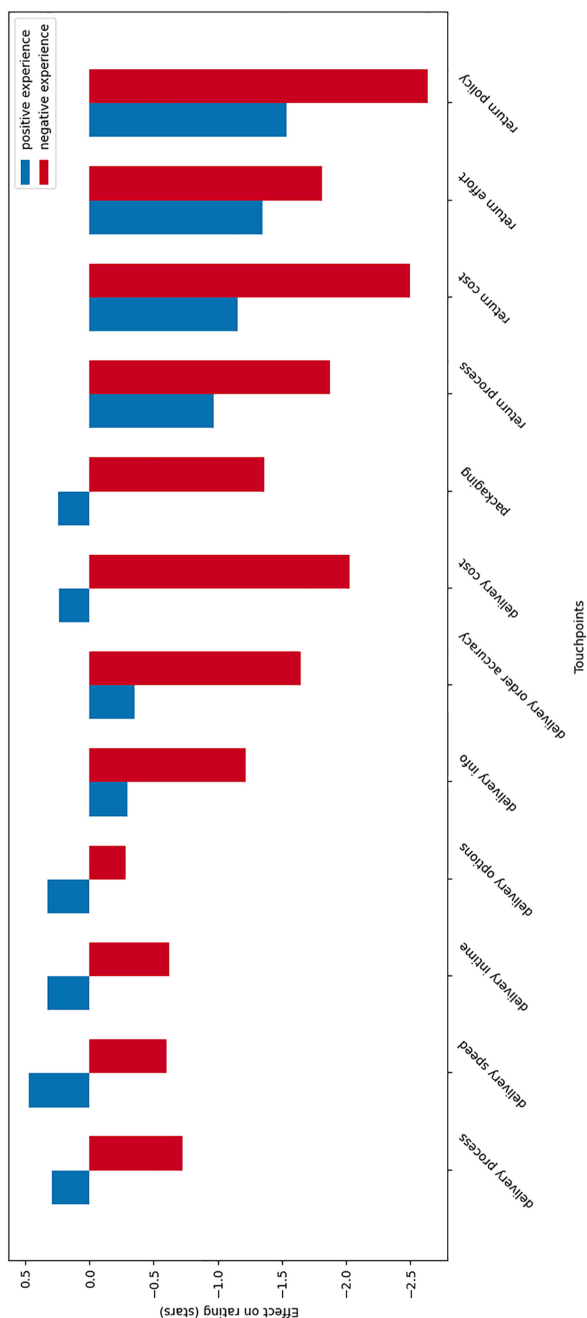
average five-star ratings for positive experiences and a significantly higher number of one-star ratings for negative experiences (see [Figure 2: e and f](#)). This pattern suggests a critical consumer emphasis on accurate and transparent delivery information, as well as the precision of order fulfillment. The contrast in ratings, especially in one-star categories for delivery information and order accuracy during negative experiences, i.e. the gap between the average and negative experience, suggests an apparent consumer sensitivity to these specific delivery touchpoints in order fulfillment.

4.1.2 Influence of packaging touchpoint on ratings. In the order fulfillment process, packaging acts as a critical touchpoint, serving as the consumer's first physical interaction with the product and significantly shaping their perceptions of service quality and the product's intrinsic value. The fact that consumers mention their packaging experience in their reviews is itself a strong indicator of its importance. The analysis shows a clear trend where positive packaging experiences are frequently associated with an increase in ratings, particularly five-star reviews, compared to the baseline (see [Figure 2: h](#)). Conversely, negative packaging experiences tend to result in lower ratings. Traditionally, packaging has the following basic functions – containment, protection, apportionment, unitization, information communication and convenience ([Robertson, 1990](#)). Consumer comments typically reference the condition of the primary or secondary packaging, emphasizing whether the package was delivered in perfect condition or damaged. This feedback emphasizes the protective function of packaging in e-commerce, influencing consumer experience positively or negatively.

4.1.3 Influence of return touchpoints on ratings. The study identifies four return-related touchpoints: return process, return cost, return effort and return policy. The findings indicate a consistent trend toward lower ratings in reviews that mention returns, regardless of whether the return is perceived as positive or negative. Notably, any interaction with the identified return touchpoints generally correlates with a decrease in the likelihood of receiving high ratings – even in situations where the return experience was considered positive. A negative return experience makes the loss of the five-star ratings even more dramatic. Contrasting with the expected “J-shaped” distribution of ratings for products without major experience defects, the data concerning return touchpoints suggest an inversion of this pattern ([Figure 2: i, j, k and l](#)). This inversion is particularly evident in the return cost and policy touchpoints. This pattern implies that consumers who engage in returns are likely to have had a more negative experience than those who do not initiate returns, culminating in lower star ratings. It is inferred that the return touchpoints do not directly cause lower satisfaction; rather, they are correlated with underlying factors triggering a return. Hence, consumers who typically interact with return touchpoints already have a degree of dissatisfaction and a negative return experience exacerbates this dissatisfaction, further diminishing the overall star ratings.

4.1.4 Influence of order fulfillment touchpoints on consumer experience. The 12 order fulfillment touchpoints vary significantly in their impact on star ratings, indicating that each touchpoint has a distinct influence on the overall consumer experience (see [Figure 3](#)). While a positive experience generally leads to a better rating than a negative one, the specific degree of influence each touchpoint has differs significantly. For instance, while a negative packaging experience can reduce the overall rating by nearly 1.5 stars on average, a positive packaging experience enhances it by approximately 0.2 stars.

Interestingly, a positive experience does not always translate to an increase in ratings. In fact, for six order fulfillment touchpoints, a positive experience correlates with a decrease in the average star rating. This counterintuitive finding can be explained by three underlying causes. Firstly, [Figure 3](#) shows average values. Since a positive experience that is already given five stars cannot increase from this maximum value. Secondly, from a consumer perspective, the act of returning a product can resonate negatively. A smooth return process can mitigate this negative experience but, contrastingly, an unsatisfactory one can



Source(s): Figure created by the authors

Figure 3.
Order fulfillment
touchpoints' effect on
online user ratings

considerably diminish the rating. Finally, the multifaceted nature of star ratings captures diverse touchpoints. Hence, a single review might reflect a range of touchpoints, both positive and negative. This is particularly evident for touchpoints like delivery information and order accuracy, where positive experiences are often overshadowed by negative encounters with other touchpoints (see examples in Table 3).

4.2 Significance test

To determine the significance of the observed patterns, the study investigated whether the over- and underrepresentation in star ratings could be attributed to random variation. The significance test reveals a consistent pattern across all touchpoints within the 26 product categories, especially in cases with a substantial number of reviews. Examining the outcomes detailed in Appendix 2, the relative significance of touchpoints remains relatively steady across various product categories. Noteworthy positive influences on star ratings stem from on-time delivery, positive delivery speed and positive packaging experiences. On the other hand, the most noticeable negative impacts on star ratings arise from negative delivery cost, negative order inaccuracy, negative packaging experience and all return-related touchpoints. Particularly noticeable star rating declines are observed in negative delivery costs, negative return costs and the negative return process. Specific categories, including but not limited to grocery and gift cards, offer limited insights due to their smaller sample sizes.

5. Discussion

The study shows that order fulfillment touchpoints impact star ratings: a positive experience increases the rating, while a negative experience decreases it. Theories in behavioral science

Touchpoint	Product	Rating	Review
"Positive" delivery information	"Brand" Women's Trends High Waisted Double Slits Maxi Skirt	1	<i>"Shirt fit VERY BIG!! The size must be off!! I'll definitely shop again. Great price and fast service. I had tracking sent to my phone so I tracked it the whole time"</i>
"Positive" order accuracy	"Brand" Women's Grid Printed Crop Pants	2	<i>"These pants looked weird on me. The size was correct but my legs are average length (I'm 5'6 for reference) but they were too long and the shape at the hips around to the bottom was ill fitted"</i>
"Positive" return cost	"Brand" Men's Madison Cap Toe Oxford	1	<i>"The quality of this shoe is really bad. The leather feels like cardboard, the tongue is held on by some flimsy cloth, just flops around. They would be trash after the first time worn. Bargain basement shoes are better. They are going right back. Thankfully Amazon has free returns on them."</i>
"Positive" return effort	2014 "Brand" Golf Flat Bill Tour Fitted Golf Cap	2	<i>"The hat was too big. Returning was easy, but I was not able to get the size I needed after returning"</i>
"Positive" return policy	"Brand" Silent Dog Whistle	1	<i>"Absolutely useless. A waste of money. I adjusted it to the right frequency but either my dog is deaf or the think didn't work. Because he just sat there and looked at me as I blew and blew. Luckily Amazon has a great return policy"</i>

Table 3.
Positive touchpoints
encounter and negative
rating: review and
rating sample

Source(s): Table created by the authors

and psychology highlight that positive and negative events influence overall consumer perceptions and behaviors differently (Rozin and Royzman, 2001). We adopt two of these theories to discuss how positive and negative experiences of order fulfillment touchpoints may affect consumer perceptions and behaviors differently and the implications of this for e-retailers. Subsequently, we discuss the relative importance of different order fulfillment touchpoints.

5.1 Implications of positive and negative star rating

Both negativity bias and loss aversion provide theoretical support for the results and help interpret them in slightly different ways. Negativity bias means that consumers are more likely to remember and be influenced by negative experiences, which results in stronger and more durable impressions (Rozin and Royzman, 2001). Negativity bias indicates that a negative experience in any order fulfillment touchpoint has a greater influence on star ratings than a positive experience. Therefore, we conclude that it is more important to avoid negative experiences than to exceed expectations. Accordingly, e-retailers should prioritize fully understanding customer requirements to prevent negative experiences and receive better reviews and star ratings. In line with that, previous studies have shown that better reviews and ratings attract more consumers (Floyd *et al.*, 2014).

Similarly, loss aversion suggests that individuals experience more satisfaction from not losing something than from gaining something of similar value (Kahneman and Tversky, 1979). This implies that negative experiences are perceived as losses in the order fulfillment process, and the pain of these losses is more intense than the pleasure of positive experiences. This leads to a conclusion that it is vital to mitigate negative experiences to improve overall consumer satisfaction. Companies that succeed in reducing negative experiences and improving customer satisfaction are likely to have more loyal customers (Xu, 2020). This is crucial for gaining a competitive advantage, as loyal customers often contribute to greater profit and company value (Helgesen, 2006).

5.2 Relative importance of different touchpoints for consumer experience

The relative importance of different touchpoints, viewed through the lenses of negativity bias and loss aversion, involves minimizing experiences, expressed through the red bars in Figure 3, by avoiding or mitigating negative experiences. Ideally, the experiences expressed through the red bars are eliminated while the ones represented by the blue bars (representing the effects of positive experiences) remain. The maximum improvement potential for each touchpoint is therefore the difference between the total effect on rating of the current practice and the effect on rating from negative experiences. For example, if the e-retailer manages to eliminate negative experiences related to the return cost touchpoint, many consumers may still give a negative rating simply because they need to make a return. Figure 4 presents the improvement potential of user ratings for the twelve touchpoints in our data.

All touchpoints in our study can be improved by eliminating negative experiences, but the potential varies (Figure 4). By eliminating negative experiences, five touchpoints have the potential to improve rating by more than one star: cost for deliveries and returns, packaging, delivery order accuracy and return policy. To create a well-received order fulfillment process, it seems most important for the e-retailer in this study to first improve practices that meet consumer expectations for these five touchpoints. This means that the e-retailer should have a clear and smooth return process and a thoroughly elaborated return policy, while the cost of returning items should be perceived as low as possible (Nageswaran *et al.*, 2020). Otherwise, our results indicate that consumers are likely to lower their rating. The same logic applies to delivery cost – the results indicate that it should be as low as possible to meet consumer expectations. This corresponds to Ma (2017), who concluded that longer delivery times

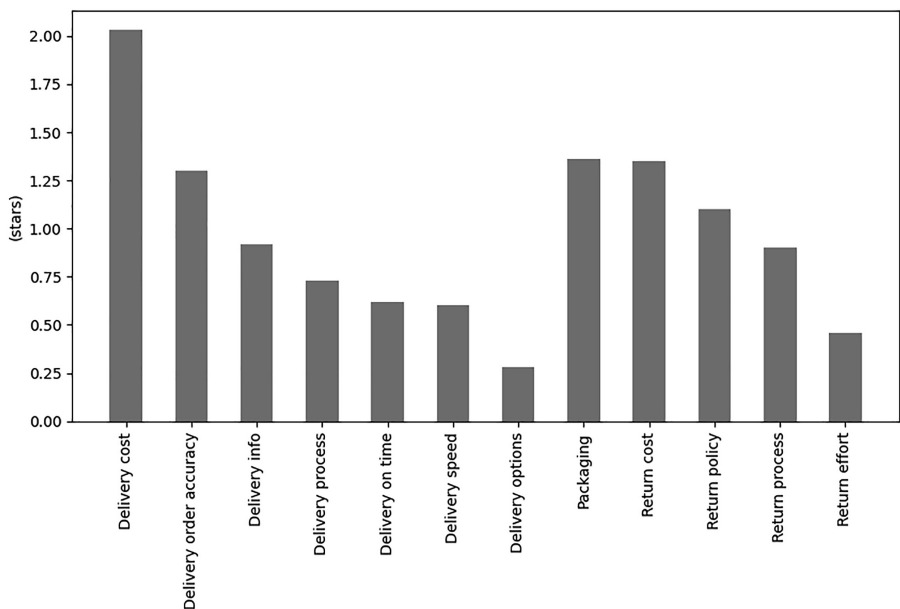


Figure 4.
Absolute difference
between the positive
and negative relative
effects of touchpoints

Source(s): Figure created by the authors

should include the shipping cost in a total price for the product, but for express shipping, a separate charge can be displayed for consumers. To ensure that packaging meets consumer expectations and to avoid a negative impact on ratings, our study showed that packaging should be undamaged. It is also important to make sure that packaging is considered sustainable or that it is not seen as excessive (Rausch *et al.*, 2021).

Notably, it may be possible to work in parallel to improve the three touchpoint categories. To achieve this, prioritization is required. For the e-retailer explored in this study, the recommended order is presented in Figure 4. Here, delivery touchpoints with greater potential can be selected to reduce delivery cost, provide accurate orders and offer clear delivery information rather than to offer alternative delivery options or faster deliveries. For returns, the most influence on star rating comes from cost, policy and process of returns. Finally, the packaging touchpoint should be addressed, as it has many negative experiences and the second most improvement potential on star rating in our study.

5.3 Consumer feedback in managing logistics

This study demonstrates that integrating consumer feedback into logistics and supply chain management is essential for enhancing e-retailer performance. Scholars have previously argued that incorporating aspects of consumer experience and feedback into the design of operations can lead to improved customer satisfaction and loyalty (Xu, 2020), enhanced brand reputation (Choi and Burnham, 2020) and more streamlined supply chain processes (Esper *et al.*, 2021). Specifically, the findings highlight that the consumer experience of the order fulfillment process contributes significantly to their overall e-retail experience. This implies that understanding and utilizing consumer feedback to guide decision-making in logistics and supply chain management can enhance overall operational effectiveness and customer satisfaction. In practice, managers who gain insights into consumer experiences,

particularly through systematic analysis of eWOM, can effectively tailor their strategies to meet customer needs. Positive consumer feedback often indicates successful practices that should be continued or expanded, while negative feedback provides invaluable insights into potential bottlenecks, operational inefficiencies and areas for improvement.

6. Conclusions

The study demonstrates key consumer journey touchpoints in e-retail order fulfillment and shows their significant impact on the overall consumer experience as well as their effect on how consumers evaluate products or services through ratings and reviews serving as a proxy for consumer satisfaction.

6.1 Managerial implications

This research offers valuable insights for managerial decision-making in e-retail and logistics. Each of the 12 touchpoints in the order fulfillment process represents an opportunity to shape the overall consumer experience, thereby influencing customer satisfaction, loyalty, and ultimately, the business's bottom line. By pinpointing these specific touchpoints, logistics and supply chain managers can concentrate more effectively on enhancing interactions directly associated with the delivery, packaging and management of product returns. In practice, this translates into a series of actionable takeaways for managers. First, identify the critical touchpoints that have the most significant impact on the consumer experience, both positively and negatively. Second, mitigating and addressing the root causes for the negative experiences in the high-impact touchpoints. Complementary to the second step, enhancing touchpoints that generate positive consumer experience to further boost customer satisfaction. To achieve a balanced and impactful improvement strategy, managers need to align cross-functional efforts and resources toward addressing the key touchpoints that cause dissatisfaction while reinforcing touchpoints resulting in positive experiences.

In addition to these operational aspects, the strategic utilization of consumer feedback, especially in the form of reviews and ratings, stands as a key implication of this study. Traditionally viewed as a means for consumer communication and trust-building, these data sources are now recognized as providing deep insights into the efficiency and effectiveness of the order fulfillment process. For example, by analyzing customer feedback using predictive models, businesses can proactively identify and address barriers or pain points in the order fulfillment journey, thus enhancing service quality and customer satisfaction. This proactive problem-solving capability to identify service design and performance issues is invaluable, offering businesses opportunities to foresee potential issues based on past feedback and take preemptive measures. This is particularly valuable for e-retailers who outsource their fulfillment operations, as it significantly enhances transparency and informs strategic decision-making. Effective utilization of consumer feedback not only assists in addressing performance gaps but also helps in tailoring services to the diverse and evolving needs of various consumer groups. For example, offering premium delivery options and packaging for high-value customers or providing a more lenient return policy for loyal customers.

6.2 Theoretical implications

This study advances theoretical understanding in both logistics and marketing by refining existing frameworks and introducing new dimensions to service quality from a consumer-centric perspective. It extends existing Logistics Service Quality (LSQ) and Physical Distribution Service Quality (PDSQ) frameworks (Mentzer *et al.*, 1989, 2001) by validating established measurements of logistics service quality in relation to fulfillment service (Rao

et al., 2011a; Xing *et al.*, 2010). Furthermore, it broadens the theoretical understanding of customer service in e-fulfillment and order fulfillment (Bressolles and Lang, 2019; Titiyal *et al.*, 2019). The established LSQ and PDSQ models, which traditionally focus on complex constructs like service convenience, quality, customer satisfaction and loyalty, are now being re-examined due to changes in retail logistics driven by digitalization, sustainable development and a shift toward consumer-centricity. This study addresses this need by identifying critical order fulfillment touchpoints that have been overlooked in traditional LSQ and PDSQ measurements (e.g. packaging experience) and demonstrating their quantifiable impact on consumer experience. Essentially, the study expands previous frameworks to include contemporary consumer experiences and preferences, which have significantly evolved in recent years, along with the analytical data techniques used for measurement.

The findings foster an integration of marketing and logistics perspectives, underscoring the significance of eWOM for service design and operations management. This is corroborated by the statistically significant volume of data analyzed, allowing reliable assumptions to be drawn about the order fulfillment service quality. Moreover, the results offer a systemic text mining method for the measurement and quantification of the effect of order fulfillment on online reviews and ratings. With that, the study offers significant implications for the fields of marketing and retail by providing novel, quantifiable data on critical touchpoints in the customer journey and highlights the role of eWOM as a proxy for measuring touchpoints. It reinterprets product ratings and reviews as comprehensive indicators of consumer experience rather than just sources of product value perception. The analysis of eWOM data supports and enhances marketing strategies and methods for utilizing consumer-generated data, integrating marketing and logistics perspectives.

6.3 Limitations and future research

The study design and findings feature limitations and simultaneously pave the way for further research. First, this study adopts a novel approach to mapping and quantifying customer journey touchpoints, which has not been done previously. This methodological novelty in utilizing consumer-generated data and text mining for fulfillment service performance measurement calls for further tests to advance and perfect the analytical approaches. From the findings' quality perspective, this limits the reliability of the list of the identified touchpoints and the interpretation of the results. Second, the values produced through effect measurements are not universally generalizable to other contexts, limiting the validity of the findings in a broad e-commerce context. Similarly, the list of touchpoints is potentially extended in the current experiences of Amazon consumers. We assume that the effect of the identified touchpoints on online user ratings and reviews may vary across different markets and time periods and calls for further investigations and testing. Furthermore, future research should examine how processes and policies should be designed to fulfill consumer expectations regarding individual touchpoints and their groups.

The data utilized in this study did not include consumer features, but previous studies show that different consumer groups rate services differently. We can assume that, based on previous scientific evidence (e.g. Sunder *et al.*, 2019) and the consistency of the main effect, some touchpoints' relative importance remains unchanged for different consumer groups, whereas others may differ. In our sample, delivery cost, delivery order accuracy and all aspects related to packaging and returns tend to have the highest absolute values. However, future research should examine the relationships among order fulfillment touchpoints and their relative importance for different consumer groups in detail. Similarly, our study noted different impacts of product categories on the relative importance of the order fulfillment touchpoints. Future research should examine these differences in detail.

Notes

1. Examples of word trees: root words “delivery” and “shipment” in sub-datasets product category “toys”
“Delivery” <https://the-impact-of-order-fulfilment-on-consumer-experience.readthedocs.io/en/latest/wordtrees/delivery.html>
“Shipment” <https://the-impact-of-order-fulfilment-on-consumer-experience.readthedocs.io/en/latest/wordtrees/shipment.html>

References

- Anderson, E.T. and Simester, D.I. (2014), “Reviews without a purchase: low ratings, loyal customers, and deception”, *Journal of Marketing Research*, Vol. 51 No. 3, pp. 249-269, doi: [10.1509/jmr.13.0209](https://doi.org/10.1509/jmr.13.0209).
- Andrejić, M. (2019), “Research in logistics service quality: a systematic literature review”, *Transport*, Vol. 35 No. 2, pp. 224-235, doi: [10.3846/transport.2019.11388](https://doi.org/10.3846/transport.2019.11388).
- Antons, D., Grünwald, E., Cichy, P. and Salge, T.O. (2020), “The application of text mining methods in innovation research: current state, evolution patterns, and development priorities”, *R&D Management*, Vol. 50 No. 3, pp. 329-351, doi: [10.1111/radm.12408](https://doi.org/10.1111/radm.12408).
- Bagdare, S. and Jain, R. (2013), “Measuring retail customer experience”, *International Journal of Retail and Distribution Management*, Vol. 41 No. 10, pp. 790-804, doi: [10.1108/ijrdm-08-2012-0084](https://doi.org/10.1108/ijrdm-08-2012-0084).
- Baxendale, S., Macdonald, E.K. and Wilson, H.N. (2015), “The impact of different touchpoints on brand consideration”, *Journal of Retailing*, Vol. 91 No. 2, pp. 235-253, doi: [10.1016/j.jretai.2014.12.008](https://doi.org/10.1016/j.jretai.2014.12.008).
- Bhatt, A., Patel, A., Chheda, H. and Gawande, K. (2015), “Amazon review classification and sentiment analysis”, *International Journal of Computer Science and Information Technologies*, Vol. 6 No. 6, pp. 5107-5110.
- Bishop, Y.M., Fienberg, S.E. and Holland, P.W. (2007), *Discrete Multivariate Analysis: Theory and Practice*, Springer Science & Business Media, New York.
- Bressolles, G. and Lang, G. (2019), “KPIs for performance measurement of e-fulfillment systems in multi-channel retailing”, *International Journal of Retail and Distribution Management*, Vol. 48 No. 1, pp. 35-52, doi: [10.1108/ijrdm-10-2017-0259](https://doi.org/10.1108/ijrdm-10-2017-0259).
- Cacioppo, J.T., Gardner, W.L. and Berntson, G.G. (1999), “The affect system has parallel and integrative processing components: form follows function”, *Journal of Personality and Social Psychology*, Vol. 76 No. 5, pp. 839-855, doi: [10.1037/0022-3514.76.5.839](https://doi.org/10.1037/0022-3514.76.5.839).
- Chen, Y. and Xie, J. (2008), “Online consumer review: word-of-mouth as a new element of marketing communication mix”, *Management Science*, Vol. 54 No. 3, pp. 477-491, doi: [10.1287/mnsc.1070.0810](https://doi.org/10.1287/mnsc.1070.0810).
- Chevalier, J.A. and Mayzlin, D. (2006), “The effect of word of mouth on sales: online book reviews”, *Journal of Marketing Research*, Vol. 43 No. 3, pp. 345-354, doi: [10.1509/jmkr.43.3.345](https://doi.org/10.1509/jmkr.43.3.345).
- Choi, L. and Burnham, T. (2020), “Brand reputation and customer voluntary sharing behavior: the intervening roles of self-expressive brand perceptions and status seeking”, *Journal of Product and Brand Management*, Vol. 30 No. 4, pp. 565-578, doi: [10.1108/jpbm-12-2019-2670](https://doi.org/10.1108/jpbm-12-2019-2670).
- Christopher, M. (2018), “New directions in logistics”, in *Global Logistics and Distribution Planning*, Routledge, pp. 27-38.
- Croxton, K.L. (2003), “The order fulfillment process”, *International Journal of Logistics Management*, Vol. 14 No. 1, pp. 19-32, doi: [10.1108/09574090310806512](https://doi.org/10.1108/09574090310806512).
- Dash, S.K. (2021), “Identifying and classifying attributes of packaging for customer satisfaction-a Kano model approach”, *International Journal of Production Management and Engineering*, Vol. 9 No. 1, p. 57, doi: [10.4995/ijpme.2021.13683](https://doi.org/10.4995/ijpme.2021.13683).
- De Leeuw, E.D., Hox, J. and Dillman, D. (2012), *International Handbook of Survey Methodology*, Routledge, London.

- Derczynski, L., Ritter, A., Clark, S. and Bontcheva, K. (2013), "Twitter part-of-speech tagging for all: overcoming sparse and noisy data", *The International Conference Recent Advances in Natural Language Processing*, pp. 198-206.
- Duong, Q.H., Zhou, L., Meng, M., Nguyen, T.V., Ieromonachou, P. and Nguyen, D.T. (2022), "Understanding product returns: a systematic literature review using machine learning and bibliometric analysis", *International Journal of Production Economics*, Vol. 243, 108340, doi: [10.1016/j.ijpe.2021.108340](https://doi.org/10.1016/j.ijpe.2021.108340).
- Engler, T.H., Winter, P. and Schulz, M. (2015), "Understanding online product ratings: a customer satisfaction model", *Journal of Retailing and Consumer Services*, Vol. 27, pp. 113-120, doi: [10.1016/j.jretconser.2015.07.010](https://doi.org/10.1016/j.jretconser.2015.07.010).
- Erevelles, S., Fukawa, N. and Swayne, L. (2016), "Big Data consumer analytics and the transformation of marketing", *Journal of Business Research*, Vol. 69 No. 2, pp. 897-904, doi: [10.1016/j.jbusres.2015.07.001](https://doi.org/10.1016/j.jbusres.2015.07.001).
- Esper, T.L., Castillo, V.E., Ren, K., Sodero, A., Wan, X., Croxton, K.L., Knemeyer, A.M., DeNunzio, S., Zinn, W. and Goldsby, T.J. (2021), "Everything old is new again: the age of consumer-centric supply chain management", *Journal of Business Logistics*, Vol. 41 No. 4, pp. 286-293, doi: [10.1111/jbl.12267](https://doi.org/10.1111/jbl.12267).
- Fellbaum, C. (1998), *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA.
- Festinger, L.A. (1957), *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford, CA.
- Flanagin, A.J., Metzger, M.J., Pure, R., Markov, A. and Hartsell, E. (2014), "Mitigating risk in ecommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention", *Electronic Commerce Research*, Vol. 14 No. 1, pp. 1-23, doi: [10.1007/s10660-014-9139-2](https://doi.org/10.1007/s10660-014-9139-2).
- Floyd, K., Freling, R., Alhoqail, S., Cho, H.Y. and Freling, T. (2014), "How online product reviews affect retail sales: a meta-analysis", *Journal of Retailing*, Vol. 90 No. 2, pp. 217-232, doi: [10.1016/j.jretai.2014.04.004](https://doi.org/10.1016/j.jretai.2014.04.004).
- Friedl, J.E. (2006), "Mastering regular expressions", in *Powerful Techniques for Perl and Other Tools*, O'Reilly Media.
- Gandomi, A. and Haider, M. (2015), "Beyond the hype: big data concepts, methods, and analytics", *International Journal of Information Management*, Vol. 35 No. 2, pp. 137-144, doi: [10.1016/j.ijinfomgt.2014.10.007](https://doi.org/10.1016/j.ijinfomgt.2014.10.007).
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M. and Herrera, F. (2016), "Big data preprocessing: methods and prospects", *Big Data Analytics*, Vol. 1 No. 1, 9, doi: [10.1186/s41044-016-0014-0](https://doi.org/10.1186/s41044-016-0014-0).
- Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanagan, J. and Smith, N.A. (2010), "Part-of-speech tagging for twitter: annotation, features, and experiments".
- Goh, T.T., Jamaludin, N.A.A., Mohamed, H., Ismail, M.N. and Chua, H.S. (2022), "A comparative study on part-of-speech taggers' performance on examination questions classification according to bloom's taxonomy", *Journal of Physics: Conference Series*, Vol. 2224 No. 1, 012001, doi: [10.1088/1742-6596/2224/1/012001](https://doi.org/10.1088/1742-6596/2224/1/012001).
- Griffis, S.E., Rao, S., Goldsby, T.J. and Niranjan, T.T. (2012), "The customer consequences of returns in online retailing: an empirical analysis", *Journal of Operations Management*, Vol. 30 No. 4, pp. 282-294, doi: [10.1016/j.jom.2012.02.002](https://doi.org/10.1016/j.jom.2012.02.002).
- Hansen, K.K. and van der Goot, R. (2023), "Cross-domain evaluation of POS taggers: from Wall Street journal to fandom wiki", *arXiv Preprint*, arXiv:2304.13989.
- Helgesen, Ø. (2006), "Are loyal customers profitable? Customer satisfaction, customer (action) loyalty and customer profitability at the individual level", *Journal of Marketing Management*, Vol. 22 Nos 3-4, pp. 245-266, doi: [10.1362/026725706776861226](https://doi.org/10.1362/026725706776861226).
- Hernández-Ortega, B. (2018), "Don't believe strangers: online consumer reviews and the role of social psychological distance", *Information and Management*, Vol. 55 No. 1, pp. 31-50, doi: [10.1016/j.im.2017.03.007](https://doi.org/10.1016/j.im.2017.03.007).

- Hoffart, J.C., Olschewski, S. and Rieskamp, J. (2019), "Reaching for the star ratings: a Bayesian-inspired account of how people use consumer ratings", *Journal of Economic Psychology*, Vol. 72, pp. 99-116, doi: [10.1016/j.joep.2019.02.008](https://doi.org/10.1016/j.joep.2019.02.008).
- Hu, N., Zhang, J. and Pavlou, P.A. (2009), "Overcoming the J-shaped distribution of product reviews", *Communications of the ACM*, Vol. 52 No. 10, pp. 144-147, doi: [10.1145/1562764.1562800](https://doi.org/10.1145/1562764.1562800).
- Hu, N., Bose, I., Koh, N.S. and Liu, L. (2012), "Manipulation of online reviews: an analysis of ratings, readability, and sentiments", *Decision Support Systems*, Vol. 52 No. 3, pp. 674-684, doi: [10.1016/j.dss.2011.11.002](https://doi.org/10.1016/j.dss.2011.11.002).
- Janakiraman, N., Syrdal, H.A. and Freling, R. (2016), "The effect of return policy leniency on consumer purchase and return decisions: a meta-analytic review", *Journal of Retailing*, Vol. 92 No. 2, pp. 226-235, doi: [10.1016/j.jretai.2015.11.002](https://doi.org/10.1016/j.jretai.2015.11.002).
- Joutsela, M., Latvala, T. and Roto, V. (2016), "Influence of packaging interaction experience on willingness to pay", *Packaging Technology and Science*, Vol. 30 No. 8, pp. 505-523, doi: [10.1002/pts.2236](https://doi.org/10.1002/pts.2236).
- Kahneman, D. and Tversky, A. (1979), "Prospect theory: an analysis of decision under risk", *Econometrica*, Vol. 47 No. 2, pp. 263-292, doi: [10.2307/1914185](https://doi.org/10.2307/1914185).
- Koh, N.S., Hu, N. and Clemons, E.K. (2010), "Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures", *Electronic Commerce Research and Applications*, Vol. 9 No. 5, pp. 374-385, doi: [10.1016/j.elerap.2010.04.001](https://doi.org/10.1016/j.elerap.2010.04.001).
- Lambert, D.M. (1992), "Developing a customer-focused logistics strategy", *International Journal of Physical Distribution and Logistics Management*, Vol. 22 No. 6, pp. 12-19, doi: [10.1108/eum00000000000417](https://doi.org/10.1108/eum00000000000417).
- Lemon, K.N. and Verhoef, P.C. (2016), "Understanding customer experience throughout the customer journey", *Journal of Marketing*, Vol. 80 No. 6, pp. 69-96, doi: [10.1509/jm.15.0420](https://doi.org/10.1509/jm.15.0420).
- Li, H., Mao, H. and Wang, J. (2021), "Part-of-Speech tagging with rule-based data preprocessing and transformer", *Electronics*, Vol. 11 No. 1, p. 56, doi: [10.3390/electronics11010056](https://doi.org/10.3390/electronics11010056).
- Ma, S. (2017), "Fast or free shipping options in online and Omni-channel retail? The mediating role of uncertainty on satisfaction and purchase intentions", *The International Journal of Logistics Management*, Vol. 28 No. 4, pp. 1099-1122, doi: [10.1108/ijlm-05-2016-0130](https://doi.org/10.1108/ijlm-05-2016-0130).
- McCrae, J.P., Rudnicka, E. and Bond, F. (2020), "English WordNet: a new open-source wordnet for English", *K Lexical News*, Vol. 28, pp. 37-44.
- Meftah, S., Semmar, N. and Sadat, F. (2018), "A neural network model for part-of-speech tagging of social media texts", *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Mentzer, J.T., Gomes, R. and Krapfel, R.E. (1989), "Physical distribution service: a fundamental marketing concept?", *Journal of the Academy of Marketing Science*, Vol. 17 No. 1, pp. 53-62, doi: [10.1177/009207038901700107](https://doi.org/10.1177/009207038901700107).
- Mentzer, J.T., Flint, D.J. and Hult, G.T.M. (2001), "Logistics service quality as a segment-customized process", *Journal of Marketing*, Vol. 65 No. 4, pp. 82-104, doi: [10.1509/jmkg.65.4.82.18390](https://doi.org/10.1509/jmkg.65.4.82.18390).
- Milo, T. and Somech, A. (2020), "Automating exploratory data analysis via machine learning", *An Overview Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*.
- Miner, G. (2012), *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Academic Press, London.
- Minnema, A., Bijmolt, T.H.A., Gensler, S. and Wiesel, T. (2016), "To keep or not to keep: effects of online customer reviews on product returns", *Journal of Retailing*, Vol. 92 No. 3, pp. 253-267, doi: [10.1016/j.jretai.2016.03.001](https://doi.org/10.1016/j.jretai.2016.03.001).
- Moy, P. and Murphy, J. (2016), "Problems and prospects in survey research", *Journalism and Mass Communication Quarterly*, Vol. 93 No. 1, pp. 16-37, doi: [10.1177/1077699016631108](https://doi.org/10.1177/1077699016631108).

- Mudambi, S.M. and Schuff, D. (2010), "What makes a helpful online review? A study of customer reviews on amazon.com", *MIS Quarterly*, Vol. 34 No. 1, pp. 185-200, doi: [10.2307/20721420](https://doi.org/10.2307/20721420).
- Nageswaran, L., Cho, S.-H. and Scheller-Wolf, A. (2020), "Consumer return policies in omnichannel operations", *Management Science*, Vol. 66 No. 12, pp. 5558-5575, doi: [10.1287/mnsc.2019.3492](https://doi.org/10.1287/mnsc.2019.3492).
- Nan, C. and Cui, W. (2016), "Overview of text visualization techniques", *Artificial Intelligence*, Vol. 1, pp. 11-40, doi: [10.2991/978-94-6239-186-4_2](https://doi.org/10.2991/978-94-6239-186-4_2).
- Ngo-Ye, T.L. and Sinha, A.P. (2014), "The influence of reviewer engagement characteristics on online review helpfulness: a text regression model", *Decision Support Systems*, Vol. 61, pp. 47-58, doi: [10.1016/j.dss.2014.01.011](https://doi.org/10.1016/j.dss.2014.01.011).
- Oliver, R.L. (1977), "Effect of expectation and disconfirmation on postexposure product evaluations: an alternative interpretation", *Journal of Applied Psychology*, Vol. 62 No. 4, pp. 480-486, doi: [10.1037/0021-9010.62.4.480](https://doi.org/10.1037/0021-9010.62.4.480).
- Pandey, R., Purohit, H., Castillo, C. and Shalin, V.L. (2022), "Modeling and mitigating human annotation errors to design efficient stream processing systems with human-in-the-loop machine learning", *International Journal of Human-Computer Studies*, Vol. 160, 102772, doi: [10.1016/j.ijhcs.2022.102772](https://doi.org/10.1016/j.ijhcs.2022.102772).
- Peeters, G. (1991), "Evaluative inference in social cognition: the roles of direct versus indirect evaluation and positive-negative asymmetry", *European Journal of Social Psychology*, Vol. 21 No. 2, pp. 131-146, doi: [10.1002/ejsp.2420210204](https://doi.org/10.1002/ejsp.2420210204).
- Rambocas, M. and Pacheco, B.G. (2018), "Online sentiment analysis in marketing research: a review", *Journal of Research in Interactive Marketing*, Vol. 12 No. 2, pp. 146-163, doi: [10.1108/jrim-05-2017-0030](https://doi.org/10.1108/jrim-05-2017-0030).
- Rao, S., Goldsby, T.J., Griffis, S.E. and Iyengar, D. (2011a), "Electronic logistics service quality (e-LSQ): its impact on the customer's purchase satisfaction and retention", *Journal of Business Logistics*, Vol. 32 No. 2, pp. 167-179, doi: [10.1111/j.2158-1592.2011.01014.x](https://doi.org/10.1111/j.2158-1592.2011.01014.x).
- Rao, S., Griffis, S.E. and Goldsby, T.J. (2011b), "Failure to deliver? Linking online order fulfillment glitches with future purchase behavior", *Journal of Operations Management*, Vol. 29 Nos 7-8, pp. 692-703, doi: [10.1016/j.jom.2011.04.001](https://doi.org/10.1016/j.jom.2011.04.001).
- Rao, S., Rabinovich, E. and Raju, D. (2014), "The role of physical distribution services as determinants of product returns in Internet retailing", *Journal of Operations Management*, Vol. 32 No. 6, pp. 295-312, doi: [10.1016/j.jom.2014.06.005](https://doi.org/10.1016/j.jom.2014.06.005).
- Rausch, T.M., Baier, D. and Wening, S. (2021), "Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes", *Journal of Retailing and Consumer Services*, Vol. 63, 102681, doi: [10.1016/j.jretconser.2021.102681](https://doi.org/10.1016/j.jretconser.2021.102681).
- Robertson, G.L. (1990), "Good and bad packaging: who decides?", *International Journal of Physical Distribution and Logistics Management*, Vol. 20 No. 8, pp. 37-40, doi: [10.1108/09600039010005575](https://doi.org/10.1108/09600039010005575).
- Rozin, P. and Royzman, E.B. (2001), "Negativity bias, negativity dominance, and contagion", *Personality and Social Psychology Review*, Vol. 5 No. 4, pp. 296-320, doi: [10.1207/s15327957pspr0504_2](https://doi.org/10.1207/s15327957pspr0504_2).
- Schaer, O., Kourentzes, N. and Fildes, R. (2019), "Demand forecasting with user-generated online information", *International Journal of Forecasting*, Vol. 35 No. 1, pp. 197-212, doi: [10.1016/j.ijforecast.2018.03.005](https://doi.org/10.1016/j.ijforecast.2018.03.005).
- Shang, G., McKie, E.C., Ferguson, M.E. and Galbreth, M.R. (2019), "Using transactions data to improve consumer returns forecasting", *Journal of Operations Management*, Vol. 66 No. 3, pp. 326-348, doi: [10.1002/joom.1071](https://doi.org/10.1002/joom.1071).
- Shapiro, B.P., Rangan, V.K. and Sviokla, J.J. (1993), "Staple yourself to an order", *Harvard Business Review*, Vol. 70 No. 4, pp. 113-122.
- Speklé, R.F. and Widener, S.K. (2017), "Challenging issues in survey research: discussion and suggestions", *Journal of Management Accounting Research*, Vol. 30 No. 2, pp. 3-21, doi: [10.2308/jmar-51860](https://doi.org/10.2308/jmar-51860).

- Spence, M. (1973), "Job market signaling", *The Quarterly Journal of Economics*, Vol. 87 No. 3, pp. 355-374, doi: [10.2307/1882010](https://doi.org/10.2307/1882010).
- Statista (2021), "Statista consumer insights", B2C E-commerce Report, ed. K. van Gelder.
- Statista (2022), "Statista consumer insights", B2C E-commerce Report, ed. S. Chevalier.
- Statista (2024), "Statista consumer insights", B2C E-commerce Report, ed. U. Bashir.
- Sunder, S., Kim, K.H. and Yorkston, E.A. (2019), "What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions", *Journal of Marketing*, Vol. 83 No. 6, pp. 93-112, doi: [10.1177/0022242919875688](https://doi.org/10.1177/0022242919875688).
- Tanasa, D. and Trousse, B. (2004), "Advanced data preprocessing for intersites web usage mining", *IEEE Intelligent Systems*, Vol. 19 No. 2, pp. 59-65, doi: [10.1109/mis.2004.1274912](https://doi.org/10.1109/mis.2004.1274912).
- Taylor, D., Brockhaus, S., Knemeyer, A.M. and Murphy, P. (2019), "Omnichannel fulfillment strategies: defining the concept and building an agenda for future inquiry", *The International Journal of Logistics Management*, Vol. 30 No. 3, pp. 863-891, doi: [10.1108/ijlm-09-2018-0223](https://doi.org/10.1108/ijlm-09-2018-0223).
- Titilal, R., Bhattacharya, S. and Thakkar, J.J. (2019), "E-fulfillment performance evaluation for an e-tailer: a DANP approach", *International Journal of Productivity and Performance Management*, Vol. 69 No. 4, pp. 741-773, doi: [10.1108/ijppm-12-2018-0459](https://doi.org/10.1108/ijppm-12-2018-0459).
- Toutanova, K., Klein, D., Manning, C.D. and Singer, Y. (2003), "Feature-rich part-of-speech tagging with a cyclic dependency network", *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*.
- Tukey, J.W. (1977), *Exploratory Data Analysis*, Vol. 2, Addison-Wesley, Reading, MA.
- Vakulenko, Y., Shams, P., Hellström, D. and Hjort, K. (2019), "Service innovation in e-commerce last mile delivery: mapping the e-customer journey", *Journal of Business Research*, Vol. 101, pp. 461-468, doi: [10.1016/j.jbusres.2019.01.016](https://doi.org/10.1016/j.jbusres.2019.01.016).
- Verhoef, P.C., Lemon, K.N., Parasuraman, A., Roggeveen, A., Tsiros, M. and Schlesinger, L.A. (2009), "Customer experience creation: determinants, dynamics and management strategies", *Journal of Retailing*, Vol. 85 No. 1, pp. 31-41, doi: [10.1016/j.jretai.2008.11.001](https://doi.org/10.1016/j.jretai.2008.11.001).
- Waller, M.A., Woolsey, D. and Seaker, R. (1995), "Reengineering order fulfillment", *The International Journal of Logistics Management*, Vol. 6 No. 2, pp. 1-10, doi: [10.1108/09574099510805305](https://doi.org/10.1108/09574099510805305).
- Wang, F., Liu, X. and Fang, E. (2015), "User reviews variance, critic reviews variance, and product sales: an exploration of customer breadth and depth effects", *Journal of Retailing*, Vol. 91 No. 3, pp. 372-389, doi: [10.1016/j.jretai.2015.04.007](https://doi.org/10.1016/j.jretai.2015.04.007).
- Wu, Y., Ngai, E.W.T., Wu, P. and Wu, C. (2020), "Fake online reviews: literature review, synthesis, and directions for future research", *Decision Support Systems*, Vol. 132, 113280, doi: [10.1016/j.dss.2020.113280](https://doi.org/10.1016/j.dss.2020.113280).
- Xing, Y., Grant, D.B., McKinnon, A.C. and Fernie, J. (2010), "Physical distribution service quality in online retailing", *International Journal of Physical Distribution and Logistics Management*, Vol. 40 No. 5, pp. 415-432, doi: [10.1108/09600031011052859](https://doi.org/10.1108/09600031011052859).
- Xu, X. (2020), "Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants", *Journal of Business Research*, Vol. 106, pp. 196-210, doi: [10.1016/j.jbusres.2018.07.022](https://doi.org/10.1016/j.jbusres.2018.07.022).
- Yu, J., Subramanian, N., Ning, K. and Edwards, D. (2015), "Product delivery service provider selection and customer satisfaction in the era of internet of things: a Chinese e-retailers' perspective", *International Journal of Production Economics*, Vol. 159, pp. 104-116, doi: [10.1016/j.ijpe.2014.09.031](https://doi.org/10.1016/j.ijpe.2014.09.031).

Corresponding author

Yulia Vakulenko can be contacted at: yulia.vakulenko@plog.lth.se

Appendix 1 MVHGD in rating analysis.

The probabilities for the rating distribution of a randomly drawn sample, without replacement, from the total review population are described by discrete multivariate hypergeometric distributions (MVHGD). In a population of m star-rated reviews, of which m_1 are one-star reviews, m_2 are two-star reviews, $\dots$, m_5 are five-star reviews, with $\sum_{i=1}^5 m_i = m$, a sample of size n will have probability mass function (Bishop *et al.*, 2007).

$$P(n_1, n_2, \dots, n_k) = \frac{\prod_{i=1}^k \binom{m_i}{n_i}}{\binom{m}{n}} \quad (1)$$

mean

$$\mu = \left(\frac{nm_1}{m}, \frac{nm_2}{m}, \dots, \frac{nm_k}{m} \right) \quad (2)$$

and standard deviation

$$\sigma^2 = \left(\frac{n(m-n)}{m^2(m-1)} m_1(m-m_1), \dots, \frac{n(m-n)}{m^2(m-1)} m_k(m-m_k) \right). \quad (3)$$

The hypergeometric test measures the statistical significance of a drawn sample having a specific number of occurrences n_j , and for the case of over/under-representation of successes in a sample, the hypergeometric p -value is the probability of randomly drawing more/less occurrences from the population in m total draws, i.e.

$$p = P(n_1, n_i > z_i \dots, n_k),$$

or

$$p = P(n_1, n_i < z_i \dots, n_k),$$

where z_i is a positive integer less than one or more $i \in k$.

product_category	n_reviews	av_rating	delivery cost		delivery info		delivery time		delivery options		delivery order accuracy		delivery process		packaging		return cost		return effort		return policy		return process			
			neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos		
Apparel	3227716	4.18	-2.12	0.12	-0.78	-0.26	-0.6	0.35	*	0.31	-1.46	-0.36	-0.73	0.32	-0.55	0.47	-1.34	0.35	-2.19	-0.8	-1.31	-0.69	-2.45	-0.8	-1.76	-0.69
	3227798	4.27	-2.18	0.35	*	-0.31	-0.76	0.31	*	0.29	-1.81	-0.32	-0.82	0.27	-0.62	0.46	-1.15	0.18	-2.85	-1.72	*	-1.36	*	-1.54	-1.84	-1.25
Automotive	1388823	4.24	*	*	*	*	0.39	*	*	-1.63	0.08	-0.77	0.32	*	0.46	-1.49	0.17	-2.57	-1.15	*	-1.69	*	-1.62	*	-1.53	*
	7493139	4.46	-1.55	-0.14	-0.29	-0.77	0.19	*	0.07	-1.82	-0.81	-1.08	-1.09	0.19	-0.72	0.33	-1.19	0.23	-2.76	-1.12	*	-0.64	*	-1.99	-1.33	-0.08
Books	18147345	4.03	-2.15	0.27	-1.41	-0.09	-0.52	0.45	*	0.4	-1.85	-0.32	-0.52	0.37	-0.55	0.6	-1.28	0.37	-2.41	-1.42	-1.91	-1.5	-2.45	-1.43	-1.84	-1.2
	717824	4.12	*	0.26	*	*	-0.42	0.32	*	*	-1.38	-0.03	-0.57	0.35	*	0.47	-1.24	0.18	-2.41	-1.62	*	-1.45	*	-1.79	*	-1.45
Electronics	135288	4.75	*	*	*	*	*	*	*	*	*	*	*	0.09	*	0.08	-2.22	-0.04	*	*	*	*	*	*	*	*
	1969959	4.36	*	0.27	*	-0.5	0.19	*	0.17	-1.73	-0.58	-1.0	0.2	-0.52	0.38	-1.63	0.08	*	*	*	*	*	*	*	*	
Health_Personal_Care	4432968	4.17	-1.81	0.42	*	-0.35	-0.55	0.36	*	0.41	-1.71	-0.57	-0.85	0.34	-0.56	0.53	-1.45	0.31	-2.85	-1.8	*	-1.87	*	-1.75	-1.67	-1.37
	5540929	4.21	-2.09	0.18	*	-0.56	-0.69	0.33	*	0.35	-1.85	-0.4	-0.83	0.33	-0.42	0.5	-1.51	0.23	-2.61	-1.94	-1.53	-1.58	*	-1.83	-2.05	-1.28
Home_Entertainment	523586	4.02	*	0.74	*	*	-0.35	0.65	*	0.59	-1.44	*	-0.23	0.56	*	0.56	-1.07	0.23	-2.35	-1.08	*	-1.5	*	-1.38	*	-1.35
	4094549	4.27	-2.04	0.15	*	-0.88	0.39	*	0.39	-1.7	-0.65	-0.7	0.3	-0.69	0.45	-1.5	0.11	-2.63	-1.71	*	-1.8	*	-1.82	-1.91	-1.26	
Lawn_Land_Garden	6129419	4.22	-1.75	0.26	*	-0.38	-0.41	0.36	*	0.48	-1.57	-0.23	-0.62	0.32	-0.63	0.49	-1.3	0.17	-2.61	-1.69	-2.15	-1.49	-2.72	-1.73	-1.64	-1.18
	311294	4.24	*	*	*	*	0.25	0.424	*	*	-1.73	*	*	0.28	*	0.43	-1.09	0.3	*	-1.22	*	*	*	*	*	*
Major_Appliances	68902	4.16	*	*	*	*	0.37	*	*	-1.35	*	*	*	0.21	*	0.53	-1.48	-0.07	*	*	*	*	*	*	*	*
	1952368	4.57	*	*	*	*	0.12	*	*	-1.85	*	-0.88	0.14	*	0.26	-1.31	0.11	*	*	*	*	*	*	*	*	-0.09
Musical_Instruments	780940	4.28	*	0.12	*	*	0.25	*	*	-1.85	*	-0.74	0.27	*	0.41	-1.28	0.16	-2.44	*	*	-1.87	*	-1.46	*	-1.26	*
	2256020	4.15	-1.55	0.26	0.02	-0.72	0.4	*	0.53	-1.82	-0.37	-0.76	0.37	-0.59	0.55	-1.39	0.27	-2.59	-1.32	*	-1.59	*	-1.86	-2.33	-0.99	
Office_Outdoors	2023648	4.27	0.06	0.06	-0.58	-0.82	0.33	*	*	-1.46	-0.2	-0.73	0.26	-0.85	0.41	-1.22	0.24	-2.38	-1.21	*	-1.39	*	-1.67	-1.76	-1.15	*
	63262	4.04	*	*	*	*	*	*	*	-1.56	*	*	0.16	*	0.59	-1.53	0.27	*	*	*	*	*	*	*	*	*
Personal_Care_Appliances	2320269	4.18	0.4	0.4	*	-0.94	0.32	*	0.33	-1.47	-0.72	-0.81	0.3	-0.41	0.5	-1.22	0.18	-2.43	-1.52	*	-1.43	*	-1.49	-2.26	-1.11	*
	195057	3.84	*	*	*	*	0.74	*	*	-1.69	*	*	0.49	*	0.79	-0.89	0.39	*	*	*	*	*	*	*	*	*
Sports	426241	4.26	-2.17	0.27	*	-0.2	-0.65	0.3	*	0.19	-1.64	-0.17	-0.69	0.24	-0.71	0.44	-1.25	0.21	-2.49	-1.5	*	-1.37	-2.78	-1.65	-1.64	-1.06
	4018188	4.27	-2.12	0.19	*	-0.23	-0.82	0.27	*	0.29	-1.73	-0.08	-0.87	0.28	-0.65	0.45	-1.54	0.25	-2.83	-1.99	*	-1.85	*	-1.97	-2.71	-1.22
Toys	4558895	4.41	-2.53	-0.01	-1.77	-0.57	-0.49	0.27	*	0.27	-1.83	-0.72	-0.76	0.21	-0.48	0.38	-1.23	0.24	-3.02	-1.42	*	-1.74	*	-1.95	-1.88	-0.44
	2395390	4.16	-2.12	0.12	*	-0.09	-0.57	0.39	*	*	-1.26	0.05	-0.64	0.33	-0.5	0.5	-1.35	0.35	-2.38	-1.07	*	-0.97	*	-1.25	-1.74	-1.01
Video	426241	4.26	-2.17	0.27	*	-0.2	-0.65	0.3	*	0.19	-1.64	-0.17	-0.69	0.24	-0.71	0.44	-1.25	0.21	-2.49	-1.5	*	-1.37	-2.78	-1.65	-1.64	-1.06
	4018188	4.27	-2.12	0.19	*	-0.23	-0.82	0.27	*	0.29	-1.73	-0.08	-0.87	0.28	-0.65	0.45	-1.54	0.25	-2.83	-1.99	*	-1.85	*	-1.97	-2.71	-1.22
Watches	4558895	4.41	-2.53	-0.01	-1.77	-0.57	-0.49	0.27	*	0.27	-1.83	-0.72	-0.76	0.21	-0.48	0.38	-1.23	0.24	-3.02	-1.42	*	-1.74	*	-1.95	-1.88	-0.44
	2395390	4.16	-2.12	0.12	*	-0.09	-0.57	0.39	*	*	-1.26	0.05	-0.64	0.33	-0.5	0.5	-1.35	0.35	-2.38	-1.07	*	-0.97	*	-1.25	-1.74	-1.01

Note(s): Cases with fewer than 50 observations are marked with “*”, which could render the findings statistically less significant

Source(s): Figure created by the authors