Shopping Hard or Hardly Shopping: Revealing Consumer Segments Using Clickstream Data

Melina Zavali, Ewelina Lacka, and Johannes de Smedt

Abstract—**The recent rise of big data analytics is transforming the apparel retailing industry. E-retailers, for example, effectively use large volumes of data generated as a result of their day-to-day business operations data to aid operations and supply chain management. Although logs of how consumers navigate through an e-commerce website are readily available in a form of clickstream data, clickstream analysis is rarely used to derive insights that can support marketing decisions, leaving it an under-researched area of study. Adding to this research stream by exploring the case of a U.K.-based fast-fashion retailer, this article reveals unique consumer segments and links them to the revenue they are capable of generating. Applying the partitioning around medoids algorithm to three random samples of 10 000 unique consumer visits to the e-commerce site of a fast-fashion retailer, six consumer segments are identified. This article shows that although the "mobile window shoppers" segment consists of the largest consumer segment, it attracts the lowest revenue. In contrast, "visitors with a purpose," although one of the smallest segments, generates the highest revenue. The findings of this article contribute to marketing research and inform practice, which can use these insights to target customer segments in a more tailored fashion.**

Index Terms—**Apparel retailing, big data, clickstream data, consumer segmentation, online purchase.**

I. Introduction

The apparel industry entered the fast-fashion era. With the contribution of fast-fashion brands, the fashion apparel industry has emerged to be one of the most influential industries in the recent past [1]. Nowadays, fast-fashion brands feature on the list of the 100 best global brands [2], along with brands such as Apple and Google. This signifies the prominence of the fast-fashion trend and underscores the importance of research in the area. Although there might be many factors contributing to such a prominent position of fast fashion in the apparel industry, the literature indicates that changes in consumer behavior, as well as improvements in fashion retailers' operations, have transformed the industry [3]. On the one hand, consumers become more fashion-conscious looking for increased variability of readily available and easily accessible fashion products [3], [4]. On the other hand, technological developments and the growing popularity of online shopping resulted in fashion retailers adopting e-commerce into their operations [5], [6]. As a result of these changes, a new "fast-fashion" trend has emerged, where one-quarter of sales are now made online [7].

While some retailers are able to monetize on the fast-fashion trend and take advantage of e-commerce, not all fast-fashion brands are equally successful. The literature alludes that the success of fast-fashion retailers is underpinned by their ability to effectively use data and data analytics to make key strategic decisions, including but not limited to marketing decisions [8]. For example, one of the most successful U.K. fast-fashion retailers, ASOS, is known to use data to identify high-value consumers and effectively allocate marketing spending [9]. Forrester Research reports that ASOS employs machine learning approaches to clarify consumers as valuable and potentially valuable based on information deriving from records of how consumers navigate through asos.com [8]. Acting upon insights obtained, ASOS is able to increase its online visit to purchase ratio [10], minimize loss, and experience revenue growth [11].

The crux to ASOS's success is an effective use of large volumes of clickstream data that are a byproduct of its e-commerce operations. Clickstream data refer to logs of how consumers navigate through a website [12], [13]. Although such data contain valuable information about consumers' online behavior [14]–[18], it is rarely used to derive insights that can inform marketing decisions [13], [19], [20]. As a result, clickstream analysis in marketing is still in its infancy [21], [22]. Indeed, ASOS is one of a few fast-fashion retailers that deploy machine learning approaches to process clickstream data to inform marketing actions. This is because clickstream data are a form of big data, and thus, it is characterized by large volumes, which are difficult to process [23], [24]. As shown in the ASOS example, however, if the appropriate data analytics methods are employed, clickstream data can derive valuable insights to support marketing activities [13], [23].

As demonstrated by ASOS, one potentially useful application of clickstream data in marketing research and practice is consumer segmentation [13], [25]. Consumer segmentation is defined as the division of consumers into groups of buyers who share distinct characteristics and behaviors that might require separate products or marketing mixes [26], [27]. Recognizing consumer heterogeneity [28], much has been written about consumer segmentation in the offline environment [29], [30], there is, however, a handful of research on consumer segmentation according to consumers' online behavior [31]. Existing studies on consumer online segmentation are limited in terms of insights

Manuscript received 27 August 2020; revised 25 November 2020 and 24 February 2021; accepted 26 March 2021. Date of publication 4 May 2021; date of current version 9 March 2023. Review of this article was arranged by Department Editor T.-M. Choi. *(Corresponding author: Ewelina Lacka.)*

Melina Zavali and Ewelina Lacka are with the Business School, University of Edinburgh, Edinburgh EH8 9JS, U.K. (e-mail: s1409968@sms.ed.ac.uk; ewelina.lacka@ed.ac.uk).

Johannes de Smedt is with the Faculty of Economics and Business, KU Leuven, 3000 Leuven, Belgium (e-mail: johannes.desmedt@ed.ac.uk).

Digital Object Identifier 10.1109/TEM.2021.3070069

they provide to consumer segments based on the visit occasion [22], shopping motivation [32], or browsing behavior [20]. While useful to inform web design, those studies do not show which consumer groups are profitable and which are *window shopping* and have no intention to purchase. Importantly, they do not inform marketing managers how to effectively target the most profitable consumer groups, thereby increasing the visit to purchase ratio. Recognizing that clickstream data contain valuable information about consumer online behavior, including website entry, browsing, and exit patterns [11] and that it can improve profitability and efficiency of marketing actions [19], [23], this article aims to fill this research void and reveal unique consumer segments based on the revenue they are capable of generating. Specifically, while utilizing a unique sample of clickstream data of a U.K.-based fast-fashion retailer, this article is set out to reveal consumer segments based on consumers' online behavior at the e-commerce website visit level and link identified segments to the revenue they are capable of generating. While not limited to website browsing patterns as in [20], for example, this article advances marketing research and informs practice on which tool to use to target identified consumer segments. Overall, this article contributes to the emerging research stream on the clickstream analysis in marketing research [21], [22], informs practice, and addresses the recent call for studies on big data analytics in the apparel industry [33].

II. Literature Review

The growing demand of fashion-conscious consumers for increased variability of fashion-forward products [1], [3], [4], along with the fashion retailers' adoption of e-commerce into their operations [5], [6] resulted in the emergence of the "fast-fashion" trend. Although fast fashion drives the apparel industry [1], growing consumer demand for fast-fashion products creates a number of issues ranging from the product design, production, and inventory management to the management of consumers' demand levels and behaviors [1], [31], [34]. Choi [1] notes that fashion retailers find it particularly difficult to satisfy mass-market product needs ensuring business profitability at the same time. This is because key strategic decisions have to be made in response to rapidly changing consumer behavior [35], which is particularly challenging in relation to consumer online behavior [3].

To respond to such changes promptly, fast-fashion retailers adopt a "sense-and-response" strategy [1], [36], where technology tools are used to collect large volumes of data, which when analyzed enable them to make timely yet informed decisions in response to changes occurring. Such data are generated as a result of technology deployment and their use can serve all kinds of industry applications [14], [15]. For example, fashion retailers are known to record and use past transaction data to forecast demand for the same or similar products, which informs operations and supply chain management [37]. The "sense-and-response" strategy can also inform marketing decision-making; insights obtained from large volumes of data can aid business profitability and inform marketing activities in an increasingly dynamic and highly demanding online marketplace [3]. The true potential of data in marketing research and practice, however, remains largely unexplored. This is particularly true in the context of fast-fashion retailing [19], [23], [38]. Thus, although operations and supply chain management benefit from insights deriving from big data, limited studies have addressed how fast-fashion retailers can use such data to inform their marketing actions [21], [22], [3].

Big data are often defined in terms of its volume—the quantity of the data; velocity—the speed of data collection and processing; veracity—reliability of data sources; value—the informational benefits of data; and variety—different types of data [49]. Indeed, fast-fashion retailers can access structured and unstructured data (see [11]), which can further be classified accordingly into four categories: transaction or business activity data, clickstream data, video data, and voice data. Among these, clickstream data are recognized as particularly valuable while making key strategic and tactical marketing decisions [11]. This is because clickstream data contain valuable information about consumer online behavior, including but not limited to website entry, browsing, and exit patterns. Importantly, it is linked to real-time events, and thus, it can reduce the level of risk, improve profitability, and efficiency of marketing actions [19], [23]. With research now advocating the advantages of clickstream data over other big data types [17], [25], "the need to put them into scrutiny for useful applications is perfectly understandable" [23].

Clickstream data encompass detailed logs of how consumers navigate through a website [13], [40]. Here, an e-commerce website serves as a data collection tool via which large volumes of data are generated [23], [41]–[44]. Although useful, relatively few studies use clickstream data to examine consumer online behavior and inform marketing actions. Thus, so far, clickstream data have been employed to examine several aspects of consumer online behavior, including information search [15], [16], intentions to continue browsing [18], [19], online decision-making and purchase intentions [10], [12], [35], and others [19], [20], [21], [45], [47]–[49]. All such research confirms that clickstream data are a powerful source of information about consumers and their online behavior [18], [47]. Notwithstanding, due to the nature of clickstream data, and particularly its large volumes, extracting useful information for decision-making is easier said than done [12], [13], [23], [40]. Indeed, clickstream data processing and analysis is a challenge that many fast-fashion retailers face [11], [13], [22]. Traditional statistical methods are deemed unsuitable to meet the requirements of big datasets, such as clickstream data [14], [41]. Data mining and machine learning approaches are more appropriate [19], although their application to marketing research is at an early stage.

Data mining is a process of deriving insights from big datasets, which involves supervised and unsupervised machine learning approaches. The supervised approaches depend on ground truth to be available for training. In contrast, unsupervised approaches aim to identify unknown patterns, behaviors, and relationships in the large, often unstructured, datasets [50], [51]. The ability to identify and extract such unknown relationships is what makes unsupervised approaches particularly popular these days [23]. A wide range of unsupervised approaches is now available. Most notably, centroid-based clustering techniques, such as big data variants of K-means and K-medoids [52], density-based approaches such as DBSCAN [53], and more recently neural network-based approaches [54], [55], have been proposed and applied in a variety of context, including clickstream data [56] as well as related areas such as system log clustering [57], and customer data [54]. Omran *et al.* [58] present an overview of widely used techniques and Vakali *et al.* [59] show their application for specific web data solutions. Among the above-mentioned unsupervised machine learning approaches, time-efficient K-means algorithm is one of the most popular [24], although its limitations need to be acknowledged including its sensitivity to outliers [60]. K-medoids address this limitation, but due to time complexity and computational efficiency, it is rarely applied [61].

A potentially useful application of K-means and K-medoids clustering techniques is consumer segmentation [19], [20], [24], [62], where consumer groups are identified based on their distinct characteristics and behaviors [29]. Clustering is the process during which the relationship between patterns, themes, or factors is identified in an unstructured dataset [50], [62]. Identified relationships may include groups or segments that share similar characteristics or behaviors, such as the visit occasion [22], shopping motivation [32], or browsing behavior [20]. Insights deriving from clustering can then be used for further processing, where the results of cluster analysis are coupled with other statistical methods [41], [50], [62]. Such data-driven insights are believed to be particularly valuable in marketing and revenue management [63], where clustering combined with traditional statistical methods is useful in the effort to maximize e-commerce revenue [23], [64]. Existing studies on consumer segmentation according to online behavior, however, lack insights into the profitability of consumer segments [22]–[24]. They seem to focus on website browsing patterns [20] and they ignore website entry and exit patterns [11]. Moreover, although consumer value has been studied by previous research, for example, research examining sales forecasting [65], advertising revenue [66], and revenue of brick-and-mortar stores [67], the true value of consumer groups' shopping for fast-fashion products has not been established.

Fast-fashion retailing is gaining importance. In the effort to retain its prominent position and ensure steady growth, fast-fashion retailers are known to employ clickstream analysis to inform supply chain and operation management [37]. Notably, limited studies have address challenges that fast-fashion retailers face from the marketing perspective [3]. Existing studies that employed clickstream analytics to marketing research derive insights into consumers' visit occasion [22], shopping motivation [32], and browsing behavior [20] among others, but they do not distinguish between valuable and potentially valuable consumers segments shopping online for fast-fashion products, nor do they inform marketing practice how to effectively target those consumer groups. Adding to the existing literature, this article aims to reveal consumer segments based on online behavior at the e-commerce website visit level and link segments identified to the revenue they are capable of generating. By doing so, this study aims to address the challenges that fast-fashion retailers face, highlighted by Choi [1], and shows how industry practice can effectively use data that are the result of their day-to-day business operations to derive useful insights and support marketing actions. Importantly, this article responds to the growing call for research to show the applications of clickstream analytics in the context of apparel retailing [19], [33].

III. Methodology

The aim of this article is addressed by utilizing clickstream data of a U.K.-based fast-fashion retailer. The retailer is the U.K.'s high street fast-fashion brand. It has well-established brick-and-mortar stores as well as an e-commerce website where consumers can shop for a full range of fast-fashion products. Adobe Analytics [68] is used to capture clickstream data generated as a result of a consumer using the retailer's e-commerce website. Adobe Analytics [69] is a web analytic tool that captures logs of how consumers access and navigate through online websites. The clickstream data used in this article represent 10% of the visits to the fast-fashion retailer website during March 2018. The dataset included $N = 1\,438\,333$ unique visits to the website. The volume of our data reflects the characteristics of big data outlined by Choi *et al.* [19].

To fulfill the research aim, we construct a unique dataset that comprises website entry, browsing, and exit patterns as well as revenue generated. The website entry patterns are reflected in the device a consumer use to access the website as well as a traffic source. We distinguish between desktop and mobile devices used to access the fast-fashion retailer's e-commerce website, which is in line with [13]. We divided traffic source in line with consumer journey mapping patterns, which indicates how consumers access the website [12], [71], [72], into seven distinct categories including; advertisement (i.e., traffic is driven by the display of ads), referrals (i.e., traffic is driven by links on other sites), shopping site (i.e., traffic is driven by links on shopping sites), internal traffic (i.e., traffic is driven by links on the retailer's website), social media (i.e., traffic is driven by links on social media), e-mail (i.e., traffic is driven by links on e-mails), and search (traffic is driven by clicking on search engine results). Measures assessing website browsing patterns have been used in previous research, albeit not altogether in the same study; different measures were selected to fit the scope and the research aim [15], [20], [46], [69], [70]. In this article, we use the following website browsing measures to achieve the research aim: past visits, total page viewed, time spent on the viewed page (in seconds), and overall website visit duration (in minutes). Finally, website exit patterns are captured by an indication of whether the consumer opened the cart during the visit and made the purchase. Revenue is also recorded. The full range of metrics is captured by Adobe Analytics and measures used in this article are provided in Table I. Those measures provide a more comprehensive picture of consumer online behavior than provided in previous research (see [20]). By including a wide range of measures, we are able to reveal consumers' profitability and inform the design of marketing activities, and therefore, address the objectives of this article. The dataset does not include any missing values. However, due to the sensitivity of clustering methods to outliers, all continuous variables are

TABLE I
MEASURES

Measure	Type of Data	Level of Measurement
Past Visits	Count	Ratio
Total Number of Pages Viewed	Count	Ratio
Time per Page (in seconds)	Continuous	Ratio
Visit Duration (in minutes)	Continuous	Ratio
Device used	Categorical	Nominal, Binary Coded: 0-desktop, 1-mobile
Traffic source	Categorical	Nominal- Eight Categories Advertisement, Referral, Shopping Site, Social Media, Internal, Email, Search,
Cart opened	Categorical	Nominal, Binary Coded: 0-No, 1-Yes
Purchase made	Categorical	Nominal, Binary Coded: 0-No, 1-Yes
Revenue	Count	Ratio

inspected for data points that differ significantly from other observations. Data points with values that exceeded the mean by ± 3 standard deviations are excluded from the analysis during the data cleaning process. After data cleaning, the sample size consisted of $N = 883\ 548$ unique visits.

Cluster analysis is employed to identify consumer segments. The cluster analysis is a form of an unsupervised machine learning approach, which allows to group data points sharing similar characteristics such that the measures belonging to one cluster or a group are more similar to each other as compared to measures belonging to different groups [40], [44], [50]. Due to the heavy computational nature of the analysis, three random samples of $N = 10\ 000$ unique visits are retained for further processing. This is in line with Choi *et al.* [19] who acknowledge the heavy computational nature of big data analytics, and thus, they suggest to use smaller samples extracted from big data while exploring the relationships between measures. In this case, the need to calculate a high number of pairwise distances between data samples and their respective cluster medoids which can quickly rise when a convergence of the algorithm is not reached quickly led to this situation. Hence, a parsimonious solution was followed where the sample size was established based on the changes in results that were reported by adding more samples incrementally in three clusters simultaneously. After no changes in the results were noted in terms of cluster membership in the three samples and silhouette scores were consistent, the adequate number was found to be $N = 10\ 000$. This sample size offers a good tradeoff between the size of the sample and generalizability and also avoids any unnecessarily expensive model-building efforts [73].

To calculate similarity, a metric of dissimilarity is necessary with Euclidian distance being the most often employed [50]. This metric, however, is only applicable to numeric data. As indicated in Table I, the dataset used in this article includes both continuous and categorical measures. The Gower distance is deemed to be more appropriate, as in (1). Gower distance uses a combination of distance metrics that satisfy each of the variables, namely range-normalized Manhattan distance for continuous variables and dice distance for nominal data, which can be calculated after turning each category into a binary variable [74]. The distances are then scaled between 0 and 1 and a distance matrix is constructed that consists of the weighted sum of dissimilarities for each variable calculated with the following function:

$$d\ (i, j) = \frac{\Sigma_k\ (\delta_{ijk} * d_{ijk} * w_k)}{\Sigma_k\ (\delta_{ijk} * w_k)} \tag{1}$$

whereby $d\ (i, j)$= distance between the *i*th and *j*th data points computed considering the *k*th variable.

w_k= the weight assigned to variable *k*.

After constructing the distance matrix, the silhouette width is used as an internal validation metric. The silhouette width measures how similar observation is to its cluster compared to other clusters, with values closer to 1 indicating a good cluster composition [75].

Next, we use the partitioning around medoids (PAM) algorithm to construct the clusters. This approach is similar to k-means previously used in [12], but instead of centroids, it uses medoids that are real observations rather than averages. The following procedure is adopted.

TABLE II
DESCRIPTIVE STATISTICS OF COUNT AND CONTINUOUS VARIABLES

Variable	Mean	Median	SD	Range	Skew	Kurtosis
Past Visits	10.96	4	17.30	1-114	2.82	9.00
Total Number of Pages Viewed	8.41	4	11.06	1-68	2.45	6.71
Time per Page	19.37	13	25.64	0-192.64	2.79	10.41
Visit Duration	3.82	1.17	6.46	0-39.2	2.67	7.67
Revenue	0.51	0	4.93	0-133.96	13.00	208.67

1) Random entries are chosen as the medoids.
2) Each observation is assigned to its closest medoid with the use of the distance matrix calculated before.
3) Each observation within a cluster is examined to identify the one that has the lowest average distance from the remaining observations. If such an entry exists, it is classified as the new medoid.
4) If at least one medoid has been altered, step 2) is repeated; otherwise, the process is ended.

The medoid approach tends to be more robust to outliers and noise compared to k-means, thus it appears to be more appropriate for the current analysis [76]. This clustering-based methodology has several benefits, for example, it is less susceptible to outliers and converges quickly in general [77]. However, K-medoids and PAM still face several downsides, including their relatively expensive pairwise distance-based calculation, which is not competitive with optimized decomposition-based techniques [78]. Besides, PAM is not capable of finding nonspherical subspaces in the sample space due to the use of distance-based clustering in contrast to incremental approaches such as DBSCAN. Nevertheless, PAM provides the greatest insight into the clusters in terms of explainability due to the use of medoids, which are actual data points, and recent efforts have optimized its behavior [79]. For those reasons, it was deemed to be most suitable in the present study.

Finally, a Kruskal–Wallis test is conducted, which aims to test what is revenue each cluster is capable of generating. This is followed by various direct comparisons, performed with the use of the Dunn test to provide detail on where differences in profitability existed. To account for the multiple comparisons, the p-value is adjusted with the Bonferroni correction to control for the inflated chance of committing a Type I error. All the above analyses are conducted with R (R Core Team, 2012), with the significance level set at $p < 0.05$.

IV. FINDINGS

A. Descriptive Statistics

Descriptive statistics are calculated to summarize the measures used for the analysis. For count and continuous variables (see Table I), both dispersion and distribution are examined. As evident from the results presented in Table II, the median is lower than the mean for all measures, suggesting that the data are positively skewed. This is confirmed by the measure of skewness, which exceeds the value of 2 in all cases, as does the measure of kurtosis, indicating that a normal univariate distribution cannot be assumed [80]. Thus, nonparametric tests are preferred for the subsequent analyses since they do not require the satisfaction of distribution assumptions.

The density of the categorical variables is examined next. The majority of consumers from the sample visit the website through mobile (82%), as compared to 18% that use a desktop to access the fast-fashion retailer e-commerce website. Mobile usage surpasses desktop usage considerably and emphasizes the importance of a “mobile-first” strategy [81], [82]. The most common way of generating traffic is by clicking on search engine listing results (57%), followed by visits initiated through referral links (23%). Traffic generated by e-mail is the third most common type of traffic source (13%), while social media, shopping websites, advertisements, and internal traffic represent less than 10% of the total traffic source in our sample. During the visit, a mere 9% of visitors open a cart and 2% make a purchase. From the 9% of visitors that opened a cart, 14% make a purchase. This indicates that consumers use the fast-fashion retailer’s e-commerce site but do not make a purchase online.

B. Clustering Analysis

The desired number of clusters is validated with the silhouette width using one sample of N = 10 000 unique visits to the retailers’ website. This metric is calculated after considering divisions for up to ten clusters, with the six-cluster solution achieving the highest value of approximately 0.60 in terms of silhouette score (see Fig. 1). For robustness purposes and to verify results, we repeat this process for two other samples of $N =$ 10 000 unique visits. As shown in Fig. 1. the results are visually consistent, producing similar silhouette scores for a variety of a number of clusters used across all three samples of $N = 10$ 000 unique visits, with the six-cluster solution producing the maximum silhouette width value for all samples.

Next, the PAM algorithm is applied to the three samples of $N =$ 10 000 to examine the behaviors associated with each cluster and further verify that clusters remain consistent across samples. As

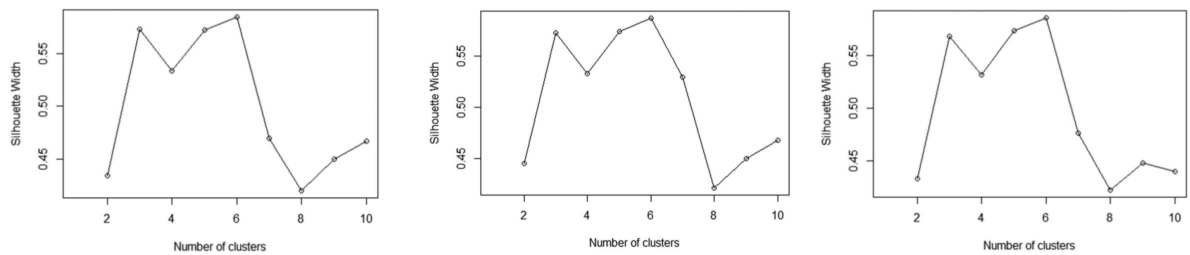

Fig. 1. Silhouette width plot from sample 1, sample 2, and sample 3.

given in Table III, all samples produce six clusters with similar behavioral characteristics and magnitude. For example, Cluster 1 is the largest in all three samples examined (i.e., Sample 1 $N = 4163$; Sample 2 $N = 4072$; Sample 3 $N = 41\ 410$), while Cluster 5 is characterized by the highest revenue in all three samples examined (i.e., Sample 1—£3.93; Sample 2—£4.93; Sample 3—£4.49).

Since results are consistent in all three samples, Sample 1 is used to describe the online behaviors of each cluster. This is consistent with [13], in which it is noted that in big data analytics, it is a common approach to determine a subset from the dataset to represent core features. For ease of interpretation, the numerical values of each cluster are classified as low, moderate, and high in accordance with their standing with respect to other clusters and the medians of the overall sample (see Table II). Except for revenue, the median is selected as the most appropriate measure of central tendency due to the skewness in the distribution of the variables previously highlighted. This results in coding where the tails only differ slightly in their coding value, e.g., for total pages viewed, 5 is already attributed a high value (H) while 25 is attributed the highest value (H^*) due to the number of samples present for each code. For the measure of revenue, the mean is assessed as the median failed to make meaningful distinctions among clusters due to the high proportion of no purchases swaying the value to zero (see Table II). Table IV presents the results of the cluster analysis.

Cluster analysis revealed six distinct clusters (i.e., consumer segments) as follows: "mobile window shopping," "enticed to buy," "examining an offer," "online window shopping," "visiting with a purpose," and "impulsive trying."'. The cluster names reflect the behavioral characteristics of consumers shopping for fast-fashion products online, and they complement goal-directed and exploratory clusters identified in [32]. Goal-directed online behavior is related to the specific or planned purchase, while exploratory behavior is less focused and no purchase is planned.

With the lowest proportion of purchases made, the "mobile window shopping" cluster reflects exploratory behavior. It is the biggest cluster with $N = 4163$ unique visits. Visits are generated by the search channel (99.3%), and this consumer group accesses the website via a mobile device (100%). Furthermore, a moderate number of previous visits (3), number of pages viewed (4), and visit duration (1.28') are typical, as well as a significant amount of time spent per page viewed (14.50"). This cluster is characterized by a lack of cart openings (0%) and the lowest proportion of purchases made (0.07%), generating the lowest revenue out of all clusters (£0.01).

In contrast, the "enticed to buy" cluster reflects goal-directed behavior. With $N = 2376$ visits is the second largest cluster generating the highest revenue. These consumers use mobile to access the website (85.9%). The traffic to the website is driven by links found on referrers' sites (97.98%). The cluster is characterized by a high number of past visits (6), a low number of pages viewed (3), and time spent per page (6.14'), as well as the lowest visit duration (0.82'), all of which indicate that those consumers superficially explore the website. Despite the short duration of the visit, a relatively high proportion of consumers open a cart (9.89%) and make a purchase (2.90%), leading to high revenue (£1.23).

The third-largest cluster named "examining the offer" consists of $N = 1394$ unique visits. These visits are made through a mobile device (92.40%) and traffic to the website is driven by e-mail marketing (96.56%). These consumers are characterized by the highest number of past visits (8), which comes as no surprise since the traffic source suggests that consumers receive marketing e-mail from the retailer (96.56%). This consumer group is characterized by the lowest number of pages viewed (2) and a moderate time spent per page (11.35"). The visits to the fast-fashion retailer website are short (0.82'). Moreover, although a moderate proportion of consumers open a cart during the visit (7.60%), a low percentage makes a purchase (0.50%) resulting in low revenue (£0.10). Aligning with previous research [32], this implies that consumers perform exploratory searches and have no purchase intentions.

The "online window shopping" cluster, a group consisting of $N = 1226$ unique visits, is characterized by a low number of prior visits (2), but a high number of total pages viewed (5). Those consumers also spend a significant amount of time per page (15.23") and overall during the visit to e-commerce website (1.49'). This is the only consumer group that accesses the website of the fast-fashion retailer using a desktop (100%). A very small group of "online window shopping" consumers opens a cart (1.47%) and makes a purchase (0.16%), resulting in low revenue (£0.05). The "online window shopping" segment, therefore, also exhibits exploratory behavior [32].

The "visiting with a purpose" cluster is one of the smallest clusters ($N = 471$ unique visits). It is, however, the most profitable out of all clusters identified (£3.93). These consumers are using search engines to look for a brand and products (93.42%) and access the website via a mobile device (78.56%). This

TABLE III
PAM RESULTS

Sample	Past Visits	Total Pages Viewed	Revenue (mean)	Time per Page	Visit Duration	Device Used	Purchase made	Cart Opened	Traffic Source	N
Cluster 1.										
1	3	4	0.01	14.50	1.28	Mobile:100%	0.07%	0.00%	Search: 99.03%	4163
2	3	4	0.02	14.44	1.23	Mobile:100%	0.05%	0.00%	Search: 99.12%	4072
3	3	4	0.01	14.44	1.18	Mobile:100%	0.07%	0.00%	Search: 99.28%	4141
Cluster 2.										
1	6	3	1.23	6.14	0.32	Mobile: 85.90%	2.90%	9.89%	Referral:97.98%	2376
2	7	2	1.57	6.14	0.32	Mobile: 86.43%	3.32%	10.34%	Referral:98.51%	2409
3	6	2	1.36	4.67	0.23	Mobile: 86.41%	3.53%	8.79%	Referral:98.28%	2435
Cluster 3.										
1	8	2	0.10	11.35	0.82	Mobile: 92.40%	0.50%	7.60%	Email: 96.56%	1394
2	7	3	0.33	11.92	0.82	Mobile: 91.48%	1.08%	6.86%	Email: 96.53%	1385
3	8	3	0.21	11.57	0.80	Mobile 92.50%	0.85%	6.42%	Email: 96.44%	1293
Cluster 4.										
1	2	5	0.05	15.23	1.49	Desktop: 100%	0.16%	1.47%	Search: 92.00%	1226
2	2	5	0.02	15.00	1.48	Desktop 100%	0.16%	1.12%	Search: 93.90%	1263
3	2	5	0.05	15.00	1.07	Desktop 100%	0.16%	1.12%	Search: 93.29%	1251
Cluster 5.										
1	3	25	3.93	21.09	9.90	Mobile: 78.56%	15.29%	100.00%	Search: 93.42%	471
2	3	31	4.79	20.76	10.11	Mobile: 73.37%	17.35%	100.00%	Search: 96.12%	490
3	3	25	4.49	20.54	9.09	Mobile: 76.51%	15.52%	100.00%	Search: 95.26%	464
Cluster 6.										
1	1	3	0.21	5.50	0.32	Mobile: 94.86%	0.54%	10.81%	Social: 79.50%	370
2	1	2	0.32	5.00	0.22	Mobile: 95.54%	1.05%	6.04%	Social: 77.17%	381
3	1	2	0.20	6.00	0.30	Mobile: 95.67%	0.96%	6.01%	Social: 74.52%	416

TABLE IV
CLUSTERING ANALYSIS RESULTS OF SAMPLE 1

Cluster Name	Past Visits	Total Pages Viewed	Revenue (mean)	Time per Page	Visit Duration	Device Used	Purchase Made	Cart Opened	Traffic source	N
Mobile Window Shopping	3 (M)	4 (M)	0.01 (L*)	14.50 (H)	1.28 (M)	Mobile: 100%	0.07% (L*)	0.00% (L*)	Search: 99.3%	4163
Enticed to Buy	6 (H)	3 (L)	1.23 (H)	6.14 (L)	0.32 (L*)	Mobile: 85.90%	2.90% (H)	9.89% (H)	Referral:97.98%	2376
Examining an Offer	8 (H*)	2 (L*)	0.10 (L)	11.35 (M)	0.82 (L)	Mobile: 92.40%	0.50% (L)	7.60% (M)	Email:96.56%	1394
Online Window Shopping	2 (L)	5 (H)	0.05 (L)	15.23 (H)	1.49 (H)	Desktop.:100%	0.16% (L)	1.47% (L)	Search: 92%	1226
Visiting with a Purpose	3 (M)	25 (H*)	3.93 (H*)	21.29 (H*)	9.90 (H*)	Mobile: 78.56%	15.29% (H*)	100% (H*)	Search: 93.42%	471
Impulsive Trying	1 (L*)	3 (L)	0.21 (M)	5.50 (L*)	0.32 (L*)	Mobile: 94.86%	0.54% (L)	10.81% (H)	Social: 79.50%	370

L: Low, H: High, M: Moderate, H*: the Highest, L*: the Lowest

cluster's consumers are characterized by high engagement with e-commerce website; we observe a moderate number of past visits (3), a high number of pages viewed (25), and time spent per page (21.29") as well as overall during the visit (9.90'). All consumers in the "visiting with a purpose" cluster open the cart (100%) and 15.29% of them make a purchase, which implies goal-directed online behavior.

Last but not the least, the "impulsive trying" cluster is the smallest consumer group with $N = 370$ unique visits. Here, consumers use a mobile device to access an e-commerce website (94.86%) being redirected to the website from social media platforms (79.50%). These consumers are characterized by very low engagement with fast-fashion retailer website, with a very low number of prior visits (1), total pages viewed per visit (3), low amount of time spent visiting the page (5.50"), and very low amount of time spent overall one the website (0.32'). Although a relatively high number of "impulsive trying" consumers opened the cart (10.81%), few consumers make a purchase (054%), which results in £0.21 medium revenue.

C. Revenue per Cluster

A Kruskal–Wallis test was performed to investigate if the revenue generated during the visit is significantly influenced by cluster membership. This nonparametric test is appropriate as it does not rely on the assumption of a normal distribution for the dependent variable (revenue). It is also useful when more than two levels exist for the independent variables, which is the case in our study (cluster membership) [83]. The Kruskal–Wallis test yields a statistically significant difference for revenue generated by each of the clusters ($\chi^2(5) = 695.34$, $p < .05$) with a mean rank of £0.05 for "online window shopping," £0.22 for "impulsive trying" £0.01 for "mobile window shopping," £0.10 for "examining an offer," £1.24 for "enticed to buy," and £3.94 for "visiting with a purpose." The results confirm that significant differences exist between clusters with regard to their associated revenue.

Following the Kruskal–Wallis, a post hoc analysis is conducted to identify which clusters are distinctive in terms of revenue generated. Specifically, the Dunn test is used to compare clusters, a nonparametric test that is appropriate for the current groups as they have an unequal number of observations [84]. Results are presented in Table V.

The Dunn test reveals that the "enticed to buy" cluster is significantly different in terms of revenue generated in comparison to all other clusters and the same was the case for "visiting with a purpose." However, the "online window shopping" cluster, as well as "mobile window shopping," "impulsive trying,"' and 'examining an offer' clusters, shows no significant differences in their profitability. After considering these results, as well as the mean revenue produced by each cluster, the conclusion is reached that "visiting with a purpose" is significantly more profitable than all other clusters. "Enticed to buy" produces more revenue than "online window shopping," "mobile window shopping," "impulsive trying," and "examining an offer" but makes significantly less than "visiting with a purpose."

Overall, the cluster analysis and subsequent statistical tests reveal that although "mobile window shopping" is the biggest cluster, it is the least profitable cluster as it generates the lowest revenue out of all clusters identified. In contrast, "visiting with a purpose" cluster, although one of the smallest clusters, is more profitable than other clusters identified. the "enticed to buy" cluster generates significant revenue, but it is less profitable than the "visiting with a purpose" cluster.

TABLE V
DUNN TEST COMPARISONS

Comparison	Z	P-Value	P-Value adjusted
Online Window Shopping vs. Impulsive Trying	-0.52	0.61	1.00
Online Window Shopping vs. Mobile Window Shopping	0.23	0.82	1.00
Impulsive Trying vs. Mobile Window Shopping	0.70	0.48	1.00
Online Window Shopping vs. Enticed to Buy	-6.33	<.00	<.00*
Impulsive Trying vs. Enticed to Buy	-3.44	<.00	0.01*
Mobile Window Shopping vs. Enticed to Buy	-8.95	<.00	<.00*
Online Window Shopping vs. Visiting with a Purpose	-22.52	<.00	<.00*
Impulsive Trying vs. Visiting with a Purpose	-17.13	<.00	<.00*
Mobile Window Shopping vs. Visiting with a Purpose	-25.26	<.00	<.00*
Enticed to Buy vs. Visiting with a Purpose	-19.79	<.00	<.00*
Online Window Shopping vs. Examining an Offer	-0.69	0.49	1.00
Impulsive Trying vs. Examining an Offer	0.06	0.95	1.00
Mobile Window Shopping vs. Examining an Offer	-1.12	0.26	1.00
Enticed to Buy vs. Examining an Offer	5.79	<.00	<.00*
Visiting with a Purpose vs. Examining an Offer	22.39	<.00	<.00*

*Statistically Significant Results

V. DISCUSSION AND CONCLUSION

Fast fashion is a popular trend among fashion-conscious consumers [1]. Growing consumers' demand for fashion-forward products, as well as fashion retailers' adoption of e-commerce, accelerates it further [3], [4]. Although fast fashion drives the apparel industry [1], the management of mass-market demand for fast-fashion products presents many challenges [31], [34], including the management of consumer online behavior profitably [1]. Technological developments and advancements in data analytics can serve as a solution to challenges that fashion retailers face and inform marketing actions [23], [41]–[44]. For example, an e-commerce website is a useful data collection tool. Clickstream data generated as a result of consumers' navigating through the e-commerce website are a valuable source of information about consumer online behavior [18], [47]. When analyzed, clickstream data can derive insights and support marketing decision-making, thus maximizing e-commerce revenues [23], [69]. A useful application of clickstream data is consumer segmentation [13], where unique consumer groups or segments are identified based on distinct characteristics or behavioral patterns [24]–[26], [59]. Due to the volume of clickstream data, however, they are rarely used to support marketing decisions [13], [15]–[19]. Existing research is limited in its use of clickstream to web browsing patterns [20], providing a fragmented picture of consumer online behavior. Exploring the case of a U.K.-based fast-fashion retailer, this article addresses this research gap and responds to a growing call for research on big data analytics in apparel retailing. Specifically, this article reveals the presence of six consumer segments and links them to the revenue they generate at the website visit level. They are summarised in Table VI.

A. Theoretical Implications

This article findings derive the following implications. First, this article adds to the under-researched area of marketing; it shows the usefulness of clickstream analysis to examine consumer online behavior at the e-commerce visit level and reveals the value of consumers shopping for fast-fashion products online. Thus, so far, only a few studies have used clickstream data to analyze consumer behavior [12], [20], none of which examined the behavior of consumers shopping online for fast-fashion products. Second, this article adds to existing research on consumer segmentation based on online behavior. The findings of this article extend previous research on clustering, which

TABLE VI
CHARACTERISTICS OF THE CONSUMER SEGMENTS

Cluster name	Size	Profitability	Engagement	Number of past visits	Notes
Visiting with a Purpose	Small	High	High	High	View high number of pages/ Profitable
Enticed to Buy	Medium	Medium	Medium	High	High number of past visits/ Profitable
Examining an Offer	Medium	Low	Low	High	Respond to emails and offers/ Less profitable
Mobile Window Shopping	Large	Low	High	Mixed	Mobile-based/ Less profitable
Online Window Shopping	Large	Low	High	Mixed	Desktop-based/ Less profitable
Impulsive Trying	Small	Low	Low	Low	Open cart, do not purchase/ Potentially profitable

used clickstream data to identify "shopping types" [20] and search motivation [32]. Aligning with [32] who distinguished between goal-directed and exploratory search behavior, this article further breaks down these profiles and reveals six consumer segments; two of which exhibit goal-directed behaviors and four perform exploratory behaviors. By including website entry and exit patterns to extend the profiles of [20] with data that are captured by modern systems, i.e., the device used as well as the traffic source, this article provides a fuller picture of consumer online behavioral patterns and links them to profitability directly. It, therefore, adds to the literature and informs marketing practice. While not limited to web browsing patterns as in [20], we revealed new clusters and show that the "examining an offer" segment is channeled through e-mail, "mobile window shopping" is established by heavy mobile phone use, and "impulsive trying" is exacerbated by social media campaigns. These insights further broaden the clusters established in [20] on "hedonic browsing" and "search/deliberation," while "visiting with a purpose" still conforms with "directed buying." Using this full range of measures, this article identifies distinct consumer segments and links them to the revenue they generate, at the website visit level. This article shows that although the segment of "mobile window shoppers" consists of the largest consumer segment, it attracts the lowest revenue. In contrast, "visitors with a purpose," although one of the smallest segments, generates the highest revenue. Overall, this article contributes to the emerging research stream on big data analytics in the context of apparel retailing [33], as well as the use of clickstream data to support marketing decision-making [12], [25].

B. Limitations and Future Research

This article suffers from some limitations, which we acknowledge. First, the nature of the current study was mainly exploratory. The findings of this article derive from the exploration of a sample of clickstream data of a fast-fashion retailer based in the U.K. The focus on a single retailer is in line with previous research, which used clickstream data [12]. The advantage of using a single case study firm is that all consumers engage with the same website (i.e., the same environment conditions) during the time period under investigation. The findings may be specific to a case fast-fashion retailer examined and the western context. The research findings may not be necessarily extrapolated to other geographic locations or consumers' cultures. Future research could examine retailers operating in other than the U.K. countries (e.g., China) and see if there are any differences in consumer online behavior and, thus, segments. We encourage future research to provide an in-depth assessment of consumer segments identified. Recognizing the value of granular insights, it would be useful to get more insights into the individual segment. For example, it would be useful to include consumer demographic information. Moreover, it would be useful to assess consumer online journey and touchpoints, which lead to conversion. Considering the size of "mobile window shopping" consumer segment, future work should examine if engagement with the fast-fashion retailer online website indeed translates to purchase in-store. Finally, clickstream data are constantly evolving with more insights being available for research and practice.

REFERENCES

[1] T. M. Choi, Ed., *Information Systems for the Fashion and Apparel Industry*. New York, NY, USA: Elsevier, 2016.
[2] Accessed: Aug. 1, 2020. [Online]. Available: www.interbrand.com/best-brands/best-global-brands/2019/ranking/
[3] V. Bhardwaj and A. Fairhurst, "Fast fashion: Response to changes in the fashion industry," *Int. Rev. Retail, Distrib. Consum. Res.*, vol. 20, no. 1, pp. 165–173, 2010.
[4] T. M. Choi, N. Liu, S. C. Liu, J. Mak, and Y. T. To, "Fast fashion brand extensions: An empirical study of consumer preferences," *J. Brand Manage.*, vol. 17, pp. 472–487, 2010.

[5] L. Macchion, A. M. Morett, F. Caniato, M. Caridi, P. Danese, and A. Vinell, "International e-commerce for fashion products: What is the relationship with performance?," *Int. J. Retail Distrib. Manage.*, vol. 45, pp. 1011–1031, 2017.

[6] S. Yu, L. Hudders, and V. Cauberghe, "Are fashion consumers like schooling fish? The effectiveness of popularity cures in fashion e-commerce," *J. Bus. Res.*, vol. 85, pp. 105–116, 2018.

[7] Accessed: Aug. 1, 2020. [Online]. Available: www.forrester.com/report/Forrester+Analytics+Online+Fashion+Retail+Forecast+2017+To+2022+Global/-/E-RES145235

[8] Accessed: Nov. 1, 2020. [Online]. Available: https://go.forrester.com/blogs/machine-learning-asos-customer-value/

[9] B. P. Chamberlain, A. CArdose, C. Liu, and R. Pagliari, "Consumer lifetime value prediction using embedding," *Proc. 23rd ACM SIGKDD Int. Conf.*, vol. 23, pp. 1753–1762, 2017.

[10] Accessed: Nov. 1, 2020. [Online]. Available: https://www.statista.com/statistics/1178854/online-fashion-retailer-consumers-purchasing-in-the-uk/

[11] S. Akter and S. Wamba, "Big data analytics in e-commerce: A systematic review and agenda for future research," *Electron. Markets*, vol. 26, pp. 173–194, 2016.

[12] D. Schellong, J. Kemper, and M. Brettel, "Clickstream data as a source to uncover consumer shopping types in a longscale online setting," in *Proc. ECIS* , 2016, pp. 1–15.

[13] O. Raphaeli, A. Goldstein, and L. Fink, "Analysing online consumer behaviour in mobile and PC devices: A novel web usage mining approach," *Electron. Commerce Res. Appl.*, vol. 26, pp. 1–12, 2017.

[14] A. L. Montgomery, S. Li, S. Kannan, and J. C. Liechty, "Modeling online browsing and path analysis using clickstream data," *Market. Sci.*, vol. 23, no. 4, pp. 579–595, 2004.

[15] E. J. Johnson, W. W. Moe, P. S. Fader, S. Bellman, and G. L. Lohse, "On the depth and dynamics of online search behaviour," *Manage. Sci.*, vol. 50, pp. 299–308, 2004.

[16] W. W. Moe and P. S. Fader, "Dynamic conversion behaviour at e-commerce sites," *Market. Sci.*, vol. 50, pp. 326–335, 2004.

[17] W. Buckinx and D. Van den Poel, "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," *Eur. J. Oper. Res.*, vol. 164, pp. 252–268, 2005.

[18] R. E. Bucklin *et al.*, "Choice and the Internet: From clickstream to research stream," *Market. Lett.*, vol. 13, pp. 245–258, 2002.

[19] T. M. Choi, S. W. Wallace, and Y. Wang, "Big data analytics in operations management," *Prod. Oper. Manage.*, vol. 27, pp. 1868–1883, 2018.

[20] W. W. Moe, "Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream," *J. Consum. Psychol.*, vol. 13, no. 1/2, pp. 29–39, 2003.

[21] R. E. Bucklin and C. Sismeiro, "Click here for Internet insight: Advances in clickstream data analysis in marketing," *J. Interactive Market.*, vol. 23, no. 1, pp. 35–48. 2009.

[22] R. Olbrich and C. Holsing, "Modelling consumer purchasing behavior in social shopping communities with clickstream data," *Int. J. Electron. Commerce*, vol. 16, no. 2, pp. 15–40, 2011.

[23] T. M. Choi, H. K. Chan, and X. Yue, Recent development in big data analytics for business operations and risk management," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 81–92, Jan. 2017.

[24] Q. Su and L. Chen, "A method for discovering clusters of e-commerce interest patterns using click-stream data," *Electron. Commerce Res. Appl.*, vol. 14, no. 1, pp. 1–13, 2015.

[25] K. W. Wen and K. F. Peng, "Market segmentation via structured click-stream analysis," *Ind. Manage. Data Syst.*, vol. 102, no. 9, pp. 493–502, 2002.

[26] G. Armstrong and P. Kotler, Ed., *Marketing: An Introduction*, 11th ed. Harlow, U.K.: Pearson Education, 2013.

[27] A. Mosquera, E. J. Ayensa, C. O. Pascual, and Y. S. Murillo, "Omnichannel shopper segmentation in the fashion industry," *J. Promot. Manage.*, vol. 25, no. 5, pp. 681–699, 2019.

[28] W. W. Moe and P. S. Fader, "Dynamic conversion behavior at e-commerce sites," *Manage. Sci.*, vol. 50, no. 3, pp. 326–335, 2004.

[29] E. Ko, E. Kim, C. R. Tayor, K. H. Kim, and I. J. Kang, "Cross-national market segmentation in the fashion industry: A study of European, Korean and US consumers," *Int. Market. Rev.*, vol. 24, pp. 629–651, 2017.

[30] P. R. Cardoso, H. S. Costa, and L. A. Noviais, "Fashion consumer profiles in the Portuguese market: Involvement, innovativeness, self-expression and impulsiveness as segmentation criteria," *Int. J. Consum. Stud.*, vol. 34, pp. 638–647, 2010.

[31] G. Cachon and R. Sinney, "The value of fast fashion: Quick response, enhanced design and strategic consumer behaviour,"*Manage. Sci.*, vol. 57, pp. 779–795, 2011.

[32] C. Janiszewski, "The influence of display characteristics on visual exploratory search behaviour," *J. Consum. Res.*, vol. 25, no. 3 pp. 290–301, 1998.

[33] Accessed: Aug. 1, 2020. [Online]. Available: https://www.ieee-tems.org/call-for-papers-impacts-of-technology-management-on-the-apparel-retailing-and-manufacturing-industry-in-the-data-analytics-era-2/

[34] C. H. Lee, T. M. Choi, and T. C. E. Cheng, "Selling to strategic and loss-averse consumers: Stocking, procurement and product design policies," *Naval Res. Logistics*, vol. 62, pp. 435–453, 2015.

[35] F. Caro and V. Martinez-de-Albeniz, "Fast fashion business model overview and research opportunities," in *Retail Supply Chain Management*, N. Agrawal, S. A. Smith, Eds. New York, NY, USA: Springer, 2015, pp. 237–264.

[36] S. H. Haeckel, *Adoptive Enterprise: Creating and Leading Sense-and-Response Organisations*. Boston, MA, USA: Harvard Bus. School Press, 1999.

[37] N. Liu, S. Ren, T. M. Choi, C. L. Hui, and S. F. Ng, "Sales forecasting for fashion retailing service industry: A review," *Math. Probl. Eng.*, vol. 2013, 2013, Art. no. 738675.

[38] M. Fisher and A. Raman, "Using data and big data in retailing," *Prod. Oper. Manage.*, vol. 27, no. 90, pp. 1665–1669, 2018.

[39] H. K. Chan, X. Wang, E. Lacka, and M. Zhang, "A mixed-method approach to extracting the value of social media data," *Prod. Oper. Manage.*, vol. 25, no. 3, pp. 568–583, 2016.

[40] D. Kumar, J. C. Bezdek, M. Palaniswami, S. Rajasegarar, C. Leckie, and T. C. Havens, "A hybrid approach to clustering in big data," *IEEE Trans. Cybern.*, vol. 46, no. 10, pp. 2372–2385, Oct. 2016.

[41] U. C. Dayal and V. Narasayya, "An overview of business intelligence technology," *Commun. ACM*, vol. 54, no. 8, pp. 88–98, 2011.

[42] M. Xu, D. Gao, C. Deng, Z. Luo, and S. Sun, "Cloud computing boosts business intelligence of telecommunication industry," in *Proc. Int. Conf. Cloud Computing*, M. G. Jaatun, G. Zhao, and C. Rong, Eds. Heidelberg, Germany: Springer, 2009, pp. 224–231.

[43] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quart.*, vol. 36, no. 4, pp. 1165–1188, 2012.

[44] H. Hu, Y. Wen, T. S. Chau, and X. Li, "Towards scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014.

[45] S. Senecal, P. J. Kalczynski, and J. Nantel, "Consumers' decision-making process and their online shopping behaviour: A clickstream analysis," *J. Bus. Res.*, vol. 58, pp. 1599–1608, 2005.

[46] P. Huang, "Searching for experience on the web: An empirical examination of consumer behaviour for search and experience goods," *J. Market.*, vol. 73, pp. 55–69, 2009.

[47] R. Olbrich and C. Holsing, "Modeling consumer purchasing behaviour in social shopping communities with clickstream data," *Int. J. Electron. Commerce*, vol. 16, pp. 15–14, 2011.

[48] P. J. Danaher, "Factors affecting web site visit duration: A cross – domain analysis," *J. Market. Res.*, vol. 43, no. 2, pp. 1599–1608, 2006.

[49] Y. H. Park and P. S. Fader, "Modelling browsing behaviour at multiple websites," *Market. Sci.*, vol. 23, pp. 280–303, 2004.

[50] B. Apte, E. P. Liu, D. Pednault, and P. Smyth, "Business applications of data mining," *Commun. ACM*, vol. 45, no. 8, pp. 49–53, 2002.

[51] L. Duan and L. D. Xu, "Business intelligence for enterprise systems: A survey," *IEEE Trans. Ind. Informat.*, vol. 8, no. 3, pp. 679–687, Aug. 2012.

[52] A. Preeti and S. Varshney, "Analysis of k-means and k-medoids algorithm for big data," *Procedia Comput. Sci.*, vol. 78, pp. 507–512, 2016.

[53] E. Schubert, J. Sander, and M. Ester, "DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN," *ACM Trans. Database Syst.*, vol. 42, no. 3 pp. 1–21, 2017.

[54] K.-L. Du, "Clustering: A neural network approach," *Neural Netw.*, vol. 23, no. 1, pp. 89–107, 2010.

[55] L. Duan, and L. D. Xu, "Business intelligence for enterprise systems: A survey," *IEEE Trans. Ind. Inf.*, vol. 8. no. 3, pp. 679–687, Aug. 2012.

[56] A. Panagiotis and C. Makris, "Tsirakis n algorithms for clustering click-stream data," *Inf. Process. Lett.*, vol. 109, no. 8, pp. 381–385, 2009.

[57] M. Landeuer, F. Skopik, M. Wurzenberger, and A. Rauber, "System log clustering approaches for cyber security applications: A survey," *Comput. Secur.*, vol. 92, 2020, Art. no. 101739.

[58] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, “An overview of clustering methods,” *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007.

[59] A. Vakali, J. Pokorný, and T. Dalamagas, “An overview of web data clustering practices,” in *Proc. Int. Conf. Extending Database Technol.*, Berlin, Heidelberg, 2004, pp. 597–606.

[60] K. Voges, K. L. Pope, and M. R. Brown, “Cluster analysis of marketing data examining on-line shopping orientation: A comparison of k-means and rough clustering,” in *Heuristics and Optimization for Knowledge Discovery*. Hershey, PA, USA: Idea Group, 2002.

[61] T. Velmurugan Vand Santhanam, “Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points,” *J. Comput. Sci.*, vol. 6, no. 3, pp. 363–368 2010.

[62] A. Fahad *et al.*, “A survey of clustering algorithms for big data: Taxonomy and empiricalanalysis,” *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279, Sep. 2014.

[63] D. Smichi-Levi, “The new frontier of price optimisation,” *MIT Sloan Manage. Rev.*, vol. 59, no. 1, pp. 22–26, 2017.

[64] K. Kim, B. S. Chung, J.-Y. Jung, and J. Park, “Revenue maximizing itemset construction for online shopping services,” *Ind. Manage. Data Syst.*, vol. 113, no. 1, pp. 96–116, 2013.

[65] P. Koen, S. Leeflang, M. Teerling, and K. Eelko, “Does online information drive offline revenues?: Only for specific products and consumer segments!,” *J. Retailing*, vol. 87, no. 1, pp. 1–17, 2011.

[66] C. Currie and I. T. Rowley, “Consumer behaviour and sales forecast accuracy: What’s going on and how should revenue managers respond?,” *J. Revenue Pricing Manage.*, vol. 9, no. 4, pp. 374–376, 2010.

[67] G. Roels and K. Fridgeirsdottir, “Dynamic revenue management for online display advertising,” *J. Revenue Pricing Manage.*, vol. 8, no. 5, pp. 452–466, 2009.

[68] Accessed: Aug. 1, 2020. [Online]. Available: https://www.adobe.com/uk/

[69] M. Pagani and A. Mirabello, “The influence of personal and social-interactive engagement in social TV web sites,” *Int. J. Electron. Commerce*, vol. 16, pp. 41–68, 2011.

[70] A. Ghose, A. Goldfarb, and S. P. Han, “How is the mobile internet different? Search costs and local activities,” *Inf. Syst. Res.*, vol. 24, pp. 613–631, 2012.

[71] K. Y. Goh, J. Chu, and J. Wu, “Mobile advertising: An empirical study of temporal and spatial differences in search behaviour and advertising response,” *J. Interactive Market.*, vol. 30, pp. 34–45, 2015.

[72] M. Graff, “Individual differences in hypertext browsing strategies,” *Behav. Inf. Technol.*, vol. 24, pp. 93–99, 2005.

[73] M. Pagani and A. Mirabello, “The influence of personal and social-interactive engagement in social TV web sites,” *Int. J. Elect. Commer.*, vol. 16, pp. 41–68, 2011.

[74] J. C. Gower, “A general coefficient of similarity and some of its properties,” *Biometrics*, vol. 27, pp. 857–874, 1971.

[75] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Comput. Appl. Math.*, vol. 20. pp. 53–65. 1987.

[76] X. Jin, J. Han, and K. Medoids Clustering, *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Ed. Boston, MA, USA: Springer, 2011.

[77] W. Chih-Ping, Y. H. Lee, and C. M. Hsu, “Empirical comparison of fast partitioning-based clustering algorithms for large data sets,” *Expert Syst. Appl.*, vol. 24, no. 4, pp. 351–363, 2003.

[78] E. Papalexakis, N. D. Sidiropoulos, and R. Bro, “From k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors,” *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 493–506, Jan. 2013.

[79] H. S. Park and C. H. Jun, “A simple and fast algorithm for K-medoids clustering,” *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009.

[80] D. George and M. Mallery. *SPSS for Windows Step by Step: A Simple Guide and Reference*, 10th ed. Boston, MA, USA: Pearson, 2010.

[81] A. Y. L. Chong, “Predicting m-commerce adoption determinants: A neural network approach,” *Expert Syst. Appl.*, vol. 40, no. 2, pp. 523–530, 2013.

[82] G. Wagner, H. Schramm-Klein, and S. Steinmann, “Online retailing across e-channels and e-channel touchpoints: Empirical studies of consumer behaviour in the multichannel e-commerce environment,” *J. Bus. Res.*, vol. 107, pp. 256–270, 2020.

[83] J. H. McDonald. *Handbook of Biological Statistics*, 3rd ed. Baltimore, MD, USA: Sparky House, 2014.

[84] J. H. Zar, Ed., *Biostatistical Analysis*, 5th ed. Upper Saddle River, NJ, USA: Pearson/Prentice-Hall, 2010.

[85] A. Mukherjee and P. Nath, “Role of electronic trust in online retailing,” *Eur. J. Market.*, vol. 41, no. 9/10, pp. 1173–1202, 2007.

[86] J. Sheth, “Impact of emerging markets on marketing: Rethinking existing perspectives and practices,” *J. Market.*, vol. 75, no. 4, pp. 66. 2011.

[87] A. Ono, A. Nakamura, A. Okuno, and M. Sumikawa, “Consumer motivations in browsing online stores with mobile devices,” *Int. J. Electron. Commerce*, vol. 16, pp. 153–178. 2012.

[88] U. Sumita and J. Yoshii, “Enhancement of e-commerce via mobile accesses to the Internet,” *Electron. Commerce Res. Appl.*, vol. 9, pp. 217–227, 2010.