IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT

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

Journal:	*Transactions on Engineering Management*
Manuscript ID	TEM-20-0603.R2
Manuscript Type:	Special Section: Impacts of Technology Management on the Apparel Retailing and Manufacturing Industry in the Data Analytics Era
Keywords:	Apparel Retailing, Big Data, Clickstream Data, Consumer Segmentation, Online Purchase
Subject Category:	Digital Technologies and Analytics

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

ABSTRACT

The recent rise of big data analytics is transforming the apparel retailing industry. E-retailers, for example, effectively use large volumes of 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 UK-based fast-fashion retailer, this study 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 study 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 research contribute to marketing research and inform practice, which can use these insights to target customer segments in a more tailored fashion.

Keywords; Apparel Retailing; Big Data; Clickstream Data; Consumer Segmentation; Online Purchase

MANAGERIAL RELEVANCE STATEMENT

The findings of this research derive several managerial implications. Utilising clickstream data of a UK-based fast-fashion retailer, we reveal six unique consumers segments and link them to the revenue they are capable of generating at the e-commerce website visit level. The consumer segments provide marketing practitioners with data-driven insights into consumer characteristics based on their online behaviour. This enables industry practice to design marketing activities tailored to specific consumer groups in accordance with the online behaviours they exhibit and the revenue they are capable of generating. As a result, informed by these research findings marketing practitioners will be able to target the most profitable customer segments and thus, ensure revenue growth. The above research findings follow from the analysis of clickstream data generated as a result of consumers navigating through a fast-fashion retailer's website. This study shows how industry practice can effectively use large volumes of data that are the result of their day-to-day business operations to derive useful insights that can support marketing decision-making ensuring business profitability at the same time.

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 behaviour, 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 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 monetise 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 UK fast-fashion retailer, 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], minimise loss and experience revenue growth [11].

The crux to ASOS' success is an effective use of large volumes of clickstream data that is a by-product of its e-commerce operations. Clickstream data refers to logs of how consumers navigate through a website [12][13]. Although such data contains valuable information about consumers' online behaviour [14][15][16][17][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 is a form of big data, and thus it is characterised by large volumes, which are difficult to process [23][24]. As shown in 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 behaviours, that might require separate products or marketing mixes [26][27]. Recognising 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 behaviour [31]. Existing studies on consumer online segmentation are limited in terms of insights they provide to consumer segments based on the visit occasion [22], shopping motivation [32] or browsing behaviour [33]. 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. Recognising that clickstream data contains valuable information about consumer online behaviour, including website entry, browsing and exit patterns [11] and that it can improve profitability and efficiency of marketing actions [19][23], this research 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 UK-based fast-fashion retailer, this study is set out to reveal consumer segments based on consumers' online behaviour 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 [33] for example, this study advances marketing research and informs practice on which tool to use to target identified consumer segments. Overall, this study 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 [34].

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 behaviours [1][31][35]. 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 behaviour [36], which is particularly challenging in relation to consumer online behaviour [3].

To respond to such changes promptly, fast-fashion retailers adopt a 'sense-and-response' strategy [1][37], where technology tools are used to collect large volumes of data, which when analysed 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 [38]. 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 [39]. 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][40]. 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][39].

Big data is 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 [51]. Indeed, fast-fashion retailers can access structured and unstructured data [see 11], which can be further classified according to four categories into (i) transaction or business activity data (ii) clickstream data (iii) video data, and (iv) voice data. Among these, clickstream data is recognised as particularly valuable while making key strategic and tactical marketing decisions [11]. This is because clickstream data contains valuable information about consumer online behaviour, 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 encompasses detailed logs of how consumers navigate through a website [13][42]. Here, an e-commerce website serves as a data collection tool via which large volumes of data are generated [23][43][44][45][46]. Although useful, relatively few studies use clickstream data to examine consumer online behaviour and inform marketing actions. Thus far, clickstream data has been employed to examine several aspects of consumer online behaviour, including information search [15][16], intentions to continue browsing [18][19], online decision-making and purchase intentions [10][12][36], and others [20][47][48][19][50][51][52]. All such research confirms that clickstream data is a powerful source of information about consumers and their online behaviour [18][50]. Notwithstanding, due to the nature of clickstream data, and particularly its large volumes, extracting useful information for decision-making is easier said than done [13][23][42][53]. Indeed, clickstream data processing and analysis is a challenge 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][43]. Data mining and machine learning approaches are more appropriate [19], thought 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, behaviours and relationships in the large, often unstructured, datasets [54][55]. 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 [56], density-based approaches such as DBSCAN [57], and more recently neural network-based approaches [58][59], have been proposed and applied in a variety of context, including clickstream data [60] as well as related areas such as system log clustering [61], and customer data [61]. Omran [62] presents an overview of widely used techniques, and Vakali [63] shows 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 [64], though its limitations need to be acknowledged including its sensitivity to outliers [65]. K-medoids address this limitation, but due to time complexity and computational efficiency it is rarely applied [66].

A potentially useful application of K-means and K-medoids clustering techniques is consumer segmentation [19][20][64][67], where consumers groups are identified based on their distinct characteristics and behaviours [29]. Clustering is the process during which the relationship between patterns, themes, or factors are identified in an unstructured dataset [54][67]. Identified relationships may include groups or segments that share similar characteristics or behaviours such as the visit occasion [22], shopping motivation [32] or browsing behaviour [33]. Insights deriving from clustering can then be used for further processing, where the results of cluster analysis are coupled with other statistical methods [43][54][67]. Such data-driven insights are believed to be particularly valuable in marketing and revenue management [68], where clustering combined with traditional statistical methods is useful in the effort to maximize e-commerce revenue [23][69]. Existing studies on consumer segmentation according to online behaviour, however, lack insights into the profitability of consumer segments [22][23][24]. They seem to focus on website browsing patterns [33] 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 [70], advertising revenue [71], and revenue of brick-and-mortar stores [72], the true value of consumer groups' shopping for fast-fashion products has not been established.

Fast-fashion retailing is gaining in 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 [38]. Notably, limited studies have address challenges fast-fashion retailers face from the marketing perspective [39]. Existing studies that employed clickstream analytics to marketing research derive insights into consumers' visit occasion [22], shopping motivation [32] and browsing behaviour [33] 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 study aims to reveal consumer segments based on online behaviour 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 fast-fashion retailers face highlighted by Choi [1] and show how industry practice can effectively use data that is the result of their day-to-day business operations to derive useful insights and support marketing actions. Importantly, this study responds to the growing call for research to show the applications of clickstream analytics in the context of the apparel retailing [19][34].

III. METHODOLOGY

The aim of this study is addressed by utilizing clickstream data of a UK-based fast-fashion retailer. The retailer is the UK's high street fast-fashion brand. It has well-established brick-and-mortar stores as well as e-commerce website where consumers can shop for a full range of fast-fashion products. Adobe Analytics [73] is used to capture clickstream data generated as a result of a consumer using the retailer's e-commerce website. Adobe Analytics

[73] is a web analytic tool that captures logs of how consumers access and navigate through online websites. The clickstream data used in this study 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 [19].

To fulfil the research aim, we construct a unique dataset which comprises of 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][76][77], 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), email (i.e. traffic is driven by links on emails), 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][49][74][75]. In this study 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), 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 study are provided in Table 1.; those measures provide a more comprehensive picture of consumer online behaviour than provided in previous research [see 33]. 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 research. The dataset does not include any missing values. However, due to the sensitivity of clustering methods to outliers, all continuous variables are 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.

<Table 1 insert here>

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 [42][46][54]. 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 and others [19] who acknowledge 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 trade-off between the size of the sample and generalizability and also avoids any unnecessarily expensive model-building efforts [78].

To calculate similarity, a metric of dissimilarity is necessary with Euclidian distance being the most often employed [54]. This metric, however, is only applicable to numeric data. As indicated in Table 1, the dataset used in this study 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 [79]. 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.

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

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 [80].

Next, we use the Partitioning Around Medoids (PAM) algorithm to construct the clusters. This approach is similar to k-means previously used by [12], but instead of centroids it uses medoids which are real observations rather than averages. The following procedure was adopted: (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 [81]. This clustering-based methodology has several benefits, for example, it is less susceptible to outliers and converges quickly in general [82]. However, K-medoids and PAM still face several downsides, including their relatively expensive pairwise distance-based calculation which is not competitive with optimised decomposition-based techniques [83]. Besides, PAM is not capable of finding non-spherical subspaces in the sample space due to the use of distance-based clustering in contrast to incremental approaches such as DBSCAN. Nevertheless, as 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 optimised its behaviour [84]. 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 the 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 was 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 was 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 1) both dispersion and distribution are examined. As evident from the results presented in Table 2, the median is lower than the mean for all measures, suggesting that the data is 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 [85]. Thus, non-parametric tests are preferred for the subsequent analyses since they do not require the satisfaction of distribution assumptions.

<Table 2 insert here>

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. The mobile usage surpasses desktop usage considerably and emphasizes the importance of a “mobile-first” strategy [89][90]. 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 email 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 Figure 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 Figure 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.

<Figure 1 insert here>

Next, the PAM algorithm is applied to the three samples of N=10,000 to examine the behaviours associated with each cluster and further verify that clusters remain consistent across samples. As shown in Table 3, all samples produce six clusters with similar behavioural characteristics and magnitude. For example, Cluster 1 is the largest in all three samples examines (i.e. Sample 1 N=4163; Sample 2 N=4072; Sample 3 N= 41,410), while Cluster 5 is characterised by the highest revenue in all thee samples examined (i.e. Sample 1 - £3.93; Sample 2 - £4.93; Sample 3 - £4.49).

<Table 3 Insert here>

Since results are consistent in all three samples, Sample 1 is used to describe the online behaviours of each cluster. This is consistent with [13], who note 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 in respect to other clusters and the medians of the overall sample (see Table 2). 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 2). Table 4 presents the results of the cluster analysis.

<Table 4 Insert here>

Cluster analysis revealed six distinct clusters (i.e. consumer segments), these are 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 behavioural characteristics of consumers shopping for fast-fashion products online, and they complement goal-directed and exploratory clusters identified by [32]. Goal-directed online behaviour is related to the specific or planned purchase, while exploratory behaviour is less focused and no purchase is planned.

With the lowest proportion of purchase made, the 'Mobile Window Shopping' cluster reflects exploratory behaviour. It is the biggest cluster with N=4,163 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 characterised by lack of cart openings (0%) and the lowest proportion of purchase made (0.07%), generating the lowest revenue out of all clusters (£0.01).

In contrast, the 'Enticed to Buy' cluster reflects goal-directed behaviour. With N=2,376 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 characterised 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=1,394 unique visits. These visits are made through a mobile device (92.40%) and traffic to the website is driven by email marketing (96.56%). These consumers are characterised 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 characterised 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 search and have no purchase intentions.

The ‘Online Window Shopping’ cluster, a group consisting of N=1,226 unique visits, is characterised 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 which 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 behaviour [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 cluster's consumers are characterised 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 behaviour.

Last but not least, the 'Impulsive Trying' cluster is the smallest consumer group with N=370 unique visits. Here, consumers use a mobile device to access e-commerce website (94.86%) being redirected to the website from social media platforms (79.50%). These consumers are characterised 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 (0,54%) 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 non-parametric 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) [91]. 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 regards 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 non-parametric test that is appropriate for the current groups as they have an unequal number of observations [92]. Results are presented in Table 5 below.

<Table 5 Insert here>

The Dunn Test reveals that the 'Enticed to Buy' cluster is significantly different in terms of revenue generated in comparison to all other cluster 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 show 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.

V. DISCUSSION & 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 apparel industry [1], the management of mass-market demand for fast-fashion products presents many challenges [31][35], including the management of consumer online behaviour profitably [1]. Technological developments and advancements in data analytics can serve as a solution to challenges fashion retailers face and inform marketing actions [23][43][44][45][46]. 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 is a valuable source of information about consumer online behaviour [18][50]. When analysed clickstream data can derive insights and support marketing decision-making, thus maximize e-commerce revenues [23][93]. A useful application of clickstream data is consumer segmentation [13], where unique consumer groups or segments are identified based on distinct characteristics or behavioural patterns [25][26][64][65]. Due to the volume of clickstream data, however, is rarely used to support marketing decisions [13][15][16][17][18][19]. Existing research is limited in its use of clickstream at to web browsing patterns [33], providing a fragmented picture of consumer online behaviour. Exploring the case of a UK-based fast-fashion retailer, this study addresses this research gap and responds to a growing call for research on big data analytics in an apparel retailing. Specifically, this study reveals the presence of six consumer segments, and link them to the revenue they generate, at the website visit level. They are summarised in Table 6.

<Table 6 Insert here>

A. Theoretical Implications

This study findings derive the following implications. First, this study adds to the under-researched area of marketing; it shows the usefulness of clickstream analysis to examine consumer online behaviour at e-commerce visit level and reveals the value of consumers shopping for fast-fashion products online. Thus far, only a few studies used clickstream data to analyse consumer behaviour [12][20], none of which examined the behaviour of consumers shopping online for fast-fashion products. Second, this study adds to existing research on consumer segmentation based on online behaviour. The findings of this study extend previous research on clustering, which used clickstream data to identify 'shopping types' [20] and search motivation [32]. Aligning with [32] who distinguished between goal-directed and exploratory search behaviour, this study further breaks down these profiles and reveals six consumer segments; two of which exhibits goal-directed behaviours and four perform exploratory behaviours. By including website entry and exit patterns to extend the profiles of [33] with data which is captured by modern systems, i.e., the device used as well as the traffic source, this study provides a fuller picture of consumer online behavioural 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 [33], we revealed new clusters and show that the 'Examining an offer' segment is channelled through email, '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 [33] on 'Hedonic Browsing' and 'Search/Deliberation' while 'Visiting with a Purpose' still conforms with 'Directed Buying'. Using this full range of measures, this study identifies distinct consumer segments and links them to the revenue they generate, at the website visit level. This study 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 study contributes to the emerging research stream on big data analytics in the context of apparel retailing [34], as well as the use of clickstream data to support marketing decision-making [12][25].

B. Managerial Implications

This study provides a set of actionable managerial implications enabling marketing managers to target the most profitable consumer groups shopping for fast-fashion products online. The findings reveal that the 'Visitors with a Purpose' consumer segment is one of the smallest segments, yet it generates the highest revenue. These consumers show high engagement with a fast-fashion retailer's website. Retailers ought to increase engagement with the website, which will increase the likelihood of purchasing. In contrast to the 'Visitors with a Purpose' consumer segment, 'Mobile Window Shoppers' consists of the largest consumer segment but it attracts the lowest revenue. These visitors tend to browse the website but are unlikely to make the purchase online. It is worth noting, however, that website browsing is an inherent part of the shopping activity, and although it may not result in a purchase online it may lead to a purchase in-store [96]. Except for 'Online Window Shopping', identified consumer segments (i.e. 'Mobile Window Shopping', 'Enticed to Buy', 'Examining an Offer', 'Visiting with a Purpose' and 'Impulsive Trying') use a mobile device to access fast-fashion retailer e-commerce website, which reinforces the need for fast-fashion retailers to adopt a 'mobile-first' approach [89][90] and ensure that their e-commerce website can be accessed via a mobile device. The benefits deriving from the adoption of mobile commerce are also highlighted in the literature [97]. Finally, the traffic source illustrates that four out of six identified consumer segments visits to the website are driven by search, indicating that consumers shopping for fast fashion products are using search engines and search for keywords for which the retailers' website is optimised (e.g. retailers' brand name or a product). To ensure that consumers access e-commerce website, the fast-fashion retailer needs to ensure it is optimised for search engines such as Google. Other traffic sources include referrals, email and social media, out of which consumers redirected to the fast fashion website from the referral site appear to be the most profitable. Fast-fashion retailers therefore are encouraged to engage referrals as part of their marketing activities.

C. Limitations and future research

This study suffers from some limitations, which we acknowledge. First, the nature of the current study was mainly exploratory. The findings of this study derive from the exploration of a sample of clickstream data of a fast-fashion retailer based in the UK. 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 UK countries (e.g. China) and see if there are any differences in consumer online behaviour, and thus segments. We encourage future research to provide an in-depth assessment of consumer segments identified. Recognising 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 is 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, Elsevier, 2016.

[2] www.interbrand.com/best-brands/best-global-brands/2019/ranking/ accessed on 1st August 2020

[3] V. Bhardwaj, A. Fairhurst. Fast fashion: response to changes in the fashion industry. The International Review of Retail, Distribution and Consumer Research, 20(1), 165-173, 2010

[4] T.M. Choi, N. Liu, S.C Liu, J. Mak, Y.T. To, Fast fashion brand extensions; An empirical study of consumer preferences. Journal of Brand Management, 17, 472-487, 2010

[5] L. Macchion, A. M. Morett, F. Caniato, M. Caridi, P. Danese, A. Vinell, International e-commerce for fashion products: what is the relationship with performance? International Journal of Retail & Distribution Management. 2017

[6] S. Yu, L. Hudders, V. Cauberghe, Are fashion consumers like schooling fish? The effectiveness of popularity cures in fashion e-commerce. Journal of Business Research, 85, 105-116, 2018

[7] www.forrester.com/report/Forrester+Analytics+Online+Fashion+Retail+Forecast+2017+To+2022+Global/-/E-RES145235 accessed on 1st August 2020

[8] https://go.forrester.com/blogs/machine-learning-asos-customer-value/ accessed on 1st November 2020

[9] Chamberlain B.P., CArdose A., Liu C., Pagliari Consumer lifetime value prediction using embedding. 23rd ACM SIGKDD International Conference, 23. 2017

[10] https://www.statista.com/statistics/1178854/online-fashion-retailer-consumers-purchasing-in-the-uk/ accessed on 1st November 2020

[11] Akter S., Wamba S., Big Data analytics in e-commerce: a systematic review and agenda for future research. Electronic Markets, 26, 173-194, 2016

[12] D. Schellong, J. Kemper M Brettel. Clickstream data as a source to uncover consumer shopping types in a longscale online setting. Association for Information Systems. ECIS 2006 Proceedings. 2016

[13] O. Raphaeli, A. Goldstein, L Fink, Analysing online consumer behaviour in mobile and PC devices. A novel web usage mining approach. Electronic Commerce Research and Applications. 26, 1-12, 2017

[14] A.L. Montgomery, S. Li, S. Kannan, J.C. Liechty, Modeling Online Browsing and Path Analysis Using Clickstream Data. Marketing Science 23(4), 579-595, 2004

[15] E.J. Johnson, W.W Moe, P.S. Fader, S. Bellman, G.L. Lohse, On the depth and dynamics of online search behaviour. Management Science 50, 299–308, 2004

[16] W.W. Moe, P.S. Fader, Dynamic conversion behaviour at e-commerce sites. Marketing Science 50, 326–335, 2004

[17] W. Buckinx, D. Van den Poel, Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operations Research. 164, 252–268, 2005

[18] R.E. Bucklin, J.M. Lattin, A. Ansari, S. Gupta, D. Bell, E. Coupey, J.D. Little, C. Mela, A. Montgomery, J. Steckel, Choice and the Internet: from clickstream to research stream. Marketing Letters 13, 245–258, 2002

[19] T.M. Choi, S.W. Wallace, Y. Wang, Big data analytics in operations management, Production and Operations Management, 2018.

[20] W. W. Moe, Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream, Journal of Consumer Psychology, 13(1-2) 29–39, 2003

[21] Bucklin, R.E., Sismeiro, C. Click here for Internet insight: Advances in clickstream data analysis in marketing. Journal of Interactive Marketing, 23 (1) 35–48. 2009

[22] Olbrich R., Holsing C, Modelling Consumer Purchasing Behavior in Social Shopping Communities with Clickstream Data, International Journal of Electronic Commerce, 16(2), 15-40, 2011

[23] T.M.,Choi, H.K. Chan, X. Yue, Recent Development in Big Data Analytics for Business Operations and Risk Management. IEEE Transactions on Cybernetics. 47(1) 81-92, 2017

[24] Su, Q., Chen, L. A method for discovering clusters of e-commerce interest patterns using click-stream data. electronic commerce research and applications, 14(1), 1-13, 2015

[25] K.W.Wen, K.F. Peng, Market segmentation via structured clickstream analysis. Industrial Management & Data Systems. 102 (9). 2002

[26] G. Armstrong, P. Kotler (Ed.,) Marketing: an Introduction. Eleventh edition, Global, Harlow: Pearson Education, 2013

[27] A. Mosquera, EJ. Ayensa, C.O. Pascual, Y.S. Murillo., Omnichannel Shopper Segmentation in the Fashion Industry. Journal of Promotion Management, 25(5), 2019

[28] Moe W.W. Fader P.S, Dynamic conversion behavior at e-commerce sites

Management Science, 50 (3) (. 326-3352004

[29] E. Ko, E. Kim, C.R Tayor, K.H. Kim, I.J, Cross-national market segmentation in the fashion industry; A study of European, Korean and US consumers. International Marketing Review. 2017

[30] P.R. Cardoso, HS. Costa, L.A. Noviais, Fashion consumer profiles in the Portuguese market; involvement, innovativeness, self-expression and impulsiveness as segmentation criteria. International Journal of Consumer Studies, 2010

[31] G. Cachon, R. Sinney, The value of fast fashion; quick response, enhanced design and strategic consumer behaviour. Management Science, 57, 779-795, 2011

[32] Janiszewski, C. The influence of display characteristics on visual exploratory search behaviour. Journal of Consumer Research, 25, 3 290–301, 1998

[33] Moe, W.W. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. Journal of Consumer Psychology, 13, 1/2 , 29–40, 2003

[34] 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/ Accessed on 1st August 2020

[35] C.H. Lee, T.M. Choi, T.C.E. Cheng, Selling to strategic and loss-averse consumers: stocking, procurement and product design policies. Naval Research Logistics, 62, 435-453, 2015

[36] F. Caro, V. Martinez-de-Albeniz, Fast Fashion business model overview and research opportunities, In. Agrawal, Smith (Eds.), Retail Supply Chain Management, 237-264, 2015

[37] S.H. Haeckel. Adoptive Enterprise: Creating and Leading Sense-and-Response Organisations. Harvard Business School Press, 1999

[38] N. Liu, S. Ren, T.M. Choi, C.L. Hui, S.F. Ng, 2013. Sales Forecasting for fashion retailing service industry; a review. Mathematical Problems in Engineering, 2013

[39] Bhardwaj V., Fairhust A., Fast fashion: response to changes in the fashion industry. The International Review of Retail, Distribution and Consumer Research 20(1). 165-173, 2010

[40] M. Fisher, A. Raman. Using data and big data in retailing. Production Operation Management, 27(90), 1665-1669, 2018

[41] Chan HK., Wang X., Lacka E., Zhang M., A mixed-method approach to extracting the value of social media data. Production and Operations Management, 25(3), 568-583

[42] D. Kumar, J.C. Bezdek, M. Palaniswami, S.Rajasegarar, C.Leckie, T.C. Havens ―A hybrid approach to clustering in big data, IEEE Transactions on Cybernetics, 2015

[43] U. Chaudhuri. Dayal, V. Narasayya, An overview of business intelligence technology, Communications of the ACM, 54(8) 88–98, 2011

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

[45] H. Chen, R.H.L Chiang, V.C. Storey. Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165-1188, 2012

[46] H. Hu, Y. Wen, T.S. Chau, X.Li, Towards scalable systems for big data analytics: a technology tutorial. IEEE Access, 2, 652-687, 2014

[47] R.E. Bucklin, C. Sismeiro,. Click here for Internet insight: advances in clickstream data analysis in marketing. Journal of Interactive Marketing 23, 35–48, 2009

[48] S. Senecal, P.J. Kalczynski, J. Nantel. Consumers' decision-making process and their online shopping behaviour: a clickstream analysis. Journal of Business Research 58, 1599–1608, 2005

[49] P. Huang. Searching for experience on the web: an empirical examination of consumer behaviour for search and experience goods. Journal of Marketing 73, 55, 2009

[50] R. Olbrich, C. Holsing,. Modeling consumer purchasing behaviour in social shopping communities with clickstream data. International Journal of Electronic Commerce, 16, 15–4, 2011

[51] P.J. Danaher, Factors affecting web site visit duration: a cross – domain analysis. Journal of Marketing Research, 43, 182, 2006

[52] Y.H. Park,, P.S. Fader,. Modelling browsing behaviour at multiple websites. Marketing Science 23, 280–303, 2004

[53] D. Schellong, J., Kemper, M.Brettel. Clickstream data as a source to uncover consumer shopping types in a large scale online setting. Association for Information Systems. ECIS Proceedings, 2016

[54] B. Apte, E. P. Liu, D. Pednault, and P. Smyth, Business applications of data mining, Communications of the ACM, 45(8) 49–53, 2002

[55] L. Duan and L. D. Xu, Business intelligence for enterprise systems: A survey, IEEE Transactions on Industrial Informatics, 8(3) 679–687, . 2012.

[56] Preeti A.,, Varshney S. Analysis of k-means and k-medoids algorithm for big data."Procedia Computer Science 78, 507-512. 2016

[57] Schubert, E., Sander J., Ester Mn.DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Transactions on Database Systems (TODS) 42.3 1-21, 2017

[58] Du, K-L. Clustering: A neural network approach. Neural networks 23(1) 89-107, 2010

[59] Aljalbout E., Golkov V., Siddiqui Y., Strobel M., Cremers D., Clustering with Deep Learning; Taxonomy and New Methods. arXiv:1801.07648, 2018

[60] Panagiotis A, Makris C Tsirakis N Algorithms for clustering clickstream data." Information Processing Letters 109(8) 381-385, 2009

[61] Landeuer M., Skopik F., Wurzenberger Mn, Rauber A., System log clustering approaches for cyber security applications; A survey. Computers & Security, 92, 2020

[61] Rajagopal, Customer data clustering using data mining technique." arXiv preprint arXiv:1112.2663 , 2011

[62] Omran, M GH, Engelbrecht, A.P. Salman A. An overview of clustering methods. Intelligent Data Analysis 11(6) 583-605, 2007

[63] Vakali A, Pokorný J Dalamagas T. An overview of web data clustering practices. International Conference on Extending Database Technology. Springer, Berlin, Heidelberg, 2004.

[64] Su, Q., Chen, L. A method for discovering clusters of e-commerce interest patterns using click-stream data. electronic commerce research and applications, 14(1), 1-13, 2015

[65] Voges K., Pope K.L., Brown MR., 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, 2002

[66] Velmurugan Vand Santhanam T, Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points, Journal of Computer Science, 6(3), 363-368 2010

[67] A. Fahad, N.Alshatri, Z Tari, A. Almari, I Khalil, A.Y. Zamaya, S., Foufou, A. Bouras, A survey of clustering algorithms for big data; Taxonomy and empiricalanalysis. IEEE Transactions on emerging Topics in Computing, 2(£), 267-279, 2014

[68] D. Smichi-Levi. The new frontier of price optimisation. MIT Sloan Management Review. 59(1), 22-26, 2017

[69] K. Kim, B.S. Chung, J.-Y. Jung, and J. Park, Revenue maximizing itemset construction for online shopping services, Industrial Management & Data Systems 113(1) 96–116, 2013.

[70] Koen, P Leeflang S,. Teerling M Eelko K. Does online information drive offline revenues?: Only for specific products and consumer segments!.Journal of Retailing 87(1) 1-17, 2011

[71] Currie, C, Rowley. I.T Consumer behaviour and sales forecast accuracy: What's going on and how should revenue managers respond?. Journal of Revenue and Pricing Management 9(4) 374-376, 2010

[72] Roels, Guillaume, and Kristin Fridgeirsdottir. "Dynamic revenue management for online display advertising." Journal of Revenue and Pricing Management 8, no. 5 (2009): 452-466.

[72] Roels G, and Fridgeirsdottir.K Dynamic revenue management for online display advertising." Journal of Revenue and Pricing Management 8,(5) 452-466, 2009

[73] https://www.adobe.com/uk/ accessed on 1st August 202

[74] M. Pagani, A. Mirabello. The influence of personal and social-interactive engagement in social TV web sites. International Journal of Electronic Commerce 16, 41–68, 2011

[75] A. Ghose, A. Goldfarb, S.P. Han. How is the mobile Internet different? Search costs and local activities. Information Systems Research 24, 613–631, 2012

[76] K.Y. Goh, J. Chu, J. Wu. Mobile advertising: an empirical study of temporal and spatial differences in search behaviour and advertising response. Journal of Interactive Marketing, 30, 34–45., 2015

[77] M. Graff. Individual differences in hypertext browsing strategies. Behaviour and Information Technology. 24, 93–99, 2005

[78] Hao K., Training a single AI model can emit as much carbon as five cars in their lifetimes. MIT Technology Review, 2019

[79] J.C. Gower. A general coefficient of similarity and some of its properties. Biometrics 27, 857–874, 1971

[80] P. J. Rousseeuw. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics. 20. 53–65. 1987

[81] X Jin, J Han. K-Medoids Clustering. In: Sammut C., Webb G.I. (Ed.,) Encyclopedia of Machine Learning. Springer: Boston, MA. 2011

[82] Chih-Ping W, Lee Y.H. Hsu C/M. Empirical comparison of fast partitioning-based clustering algorithms for large data sets. Expert Systems with applications 24(4) 351-363, 2003

[83] Papalexakis, E., Sidiropoulos N.D Bro RFrom k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors. IEEE transactions on signal processing 61(2) -506, 2002

[84] Park H.S, Jun C.HA simple and fast algorithm for K-medoids clustering." Expert systems with applications 36(2)3336-3341, 2009

[85] D. George, M. Mallery. SPSS for Windows Step by Step: A Simple Guide and Reference. Tenth Edition, Boston: Pearson, 2010

[89] A.Y.L. Chong. Predicting m-commerce adoption determinants: A neural network approach. Expert Systems with Applications, 40 2) 523-530, 2013

[90] G. Wagner, H, Schramm-Klein, S. Steinmann Online retailing across e-channels and e-channel touchpoints: Empirical studies of consumer behaviour in the multichannel e-commerce environment, Journal of Business Research, 2020

[91] J.H McDonald. Handbook of Biological Statistics (3rd Ed.). Sparky House Publishing: Baltimore, Maryland, 2014

[92] J.H. Zar (Ed.,). Biostatistical Analysis. Fifth Edition, Pearson Prentice Hall: Upper Saddle River, NJ., 2010

[93] [48] K. Kim, B.S. Chung, J.-Y. Jung, and J. Park, Revenue maximizing itemset construction for online shopping services, Industrial Management & Data Systems 113(1) 96–116, 2013.

[94] A. Mukherjee, P. Nath, Role of electronic trust in online retailing. European Journal of Marketing, 41(9/10), 1173-1202, 2007

[95] J. Sheth . Impact of emerging markets on marketing: rethinking existing perspectives and practices. Journal of Marketing, 75(4), 66. 2011

[96] A. Ono, A. Nakamura, A Okuno., M Sumikawa. Consumer motivations in browsing online stores with mobile devices. International Journal of Electronic Commerce, 16,153–178. 2012

[97] Sumita U, Yoshii, J.Enhancement of e-commerce via mobile accesses to the Internet. Electronic Commerce Research and Applications 9, 217–227, 2010

Figure 1. – Silhouette Width Plot from Sample 1, Sample 2 and Sample 3

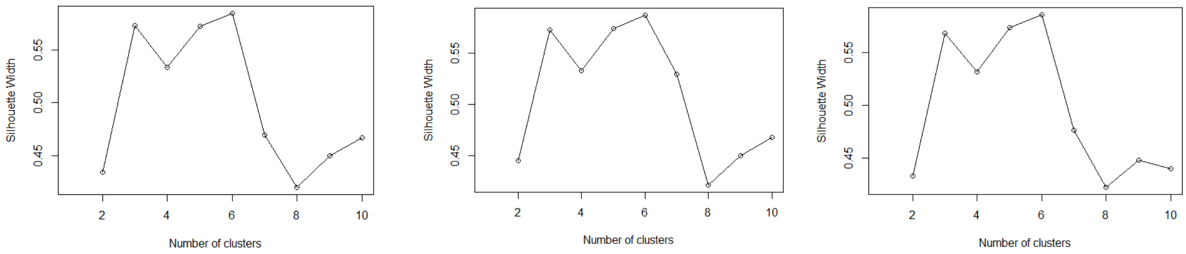

Table 1. – Measures

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

Table 2. - 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

Table 3. 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 4. 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

Table 5. - 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

Table 6. – 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

Dear Editors and Reviewers,

Thank you very much for conducting a thorough review of our manuscript. We have carefully considered the remaining issues, and further revised the manuscript. We hereby would like to submit the revision of article TEM-20-0603 to *IEEE Transactions on Engineering Management* titled **Shopping Hard or Hardly Shopping: Revealing Consumer Segments Using Clickstream Data**

Please consult the table below to see the answers to your comments, changes made in the manuscript in response to your comments/suggestions are highlighted in yellow.

Thank you

The Authors

	Comment	**Revision**
	Reviewer 1	
1.1.	The author(s) have made great effort to address my comments as well as comments raised by other reviewers. As a result, the quality of paper has been improved.	Thank you for your this positive appraisal, and the recommendation to accept the paper for the publication.
	Reviewer 2	
2.1.	Thank you for addressing my earlier comments to the paper. I strongly believe that this paper is in the right step in enriching data management in market segmentation. I wish the author(s) all the best in future research on data analytics.	Thank you for your comments and recommendation to accept the paper, we very much appreciate it.
	Reviewer 3	
3.1.	The authors have revised the manuscript substantially based on the reviewers' comments. I found it satisfactory and am pleased to see this manuscript to appear in the Transactions.	Thank you for recommending our paper for publication.

	Comment	Revision
	Reviewer 4	
4.1.	Thank you very much for revising the paper. Although the authors have responded to my comments, my concerns have not been addressed:	Thank you for reviewing the manuscript and providing further comments and suggestions. We have carefully considered your comments and revised the manuscript accordingly. We hope you will find our efforts satisfactory.
4.2.	Response 3.1. In the manuscript, the authors emphasize the research gap is "no study examining the behaviors of consumer shopping online for fast fashion". However, they have not explained how the proposed model is suitable for the fast fashion industry. In this case, are you aiming to study a generic situation or specific for fast fashion industry? Please clarify. If it is studying a generic model, then how does this model different from the existing models? Please clarify and justify the theoretical contribution and its significance.	Thank you for your comment. In this study, we examine the behaviour of consumers shopping online for fast fashion products. The specific aim of this study is '*to reveal unique consumer segments and link them to the revenue they are capable of generating*', which we highlight in the abstract. While revealing profitable and less profitable consumer groups, we also identify means by which marketing managers can effectively target these groups. We note the research gap and research objectives on p. 4-5 of the Introduction, and further reinforce it on p. 8-9 of the literature rview. To achieve this study's aim, we employ a widely used model, which is indeed generic, but which can be applied in many different settings and industries and is suitable to this research's context (i.e. to identify clusters from unstructured data). We explain this on p. 12-13 . We do not provide a state-of-the-art algorithmic extension, this is beyond the scope of this research. However, we do show how to effectively extract consumer segments and employ not used before in the context of fast fashion industry measures that allow us to use PAM to its full potential. In the discussion section see p. 20-21, we note the study contributions to marketing literature and practice. Specifically, our study provides further insights into consumer groups shopping online for fast-fashion products. We show the profitability of those consumer segments, and inform marketing practice how to target those consumers. We hope you find this response satisfactory. Thank you.

	Comment	Revision
4.3.	Response 3.2. This paper aims to segmentize the customer groups, and analyze the values generated by each group. Compared with the authors' previous study, two new measures are included in this study. However, the authors have not compared the model with and without these two measures, but jump directly to talk about the marketing aspects. In fact, the significances and impacts of these two measures haven't been examined.	Thank you for this comment. First, we would like to clarify that we did not add new measures to the revised version of the manuscript; the same set of measures have been used since the initial submission of our paper. We would like to note that there is no 'standard' set of measures used in the literature. The measures are selected based on the research aims. To that extent, Moe (2003), for example, used 'page count' as well as other webpage specific measures to study consumer search behaviour. In our paper, we also select the measures which are suitable to fulfil the objectives of our research. The unique combination of measures selected allows us to derive insights which were previously not obtained.With that in mind, we use measures providing insights into consumer web entry patterns (i.e. device used and traffic source) and web browsing (i.e. past visits, pages viewed, time spent per page, visit duration) and web exist patterns (i.e. cart opening, and purchase made), as well as revenue. By conducting cluster analysis we were able to reveal clusters which are profitable and less profitable and inform marketing practice how to effectively target those consumer groups, and thus we meet this study aim. Upon reflection, we recognize that our discussion on p.10 re. dataset construction and measures used may create confusion, and imply that there are measures commonly used in the cluster analysis of clickstream data. We have revised this section highlighting the basis upon which the measures have been selected. We hope you find this satisfactory. Thank you. Moe, W.W. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. Journal of Consumer Psychology, 13, 1/2 , 29–40, 2003