The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2177-8736.htm

Contributions to the segmentation of e-commerce nonusers: clustering the reasons not to shop online

Gabriel Nery-da-Silva
Department of Technology and Data Science, FGV-EAESP, Sao Paulo, Brazil
Marcelo Henrique de Araujo
FGV-EAESP, Sao Paulo, Brazil, and
Fernando de Souza Meirelles
Department of Technology and Data Science, FGV-EAESP, Sao Paulo, Brazil

Received 1 June 2022
Revised 1 November 2022
8 December 2022
Accepted 11 December 2022

Abstract

Purpose – The purpose is to investigate whether Brazilian e-commerce nonusers all have the same reasons not to purchase online or whether different behavior patterns might lead them to cluster in different groups.
Design/methodology/approach – This study carried out cluster analyses on a large sample (N = 9,065) from a nationwide survey on the use of information and communication technology in Brazil.
Findings – Three clusters of e-commerce nonusers were identified: the first cluster is quite reluctant; the second is characterized by disbelief in e-commerce; and the last cluster includes members who must see a product to believe it. Overall, nonusers have different reasons not to shop online, but they also share some similarities in this regard. Furthermore, socioeconomic factors do not seem to affect their behavior. The findings suggest that merchants' failure to attract customers' attention and tangibility are the major barriers to e-commerce use.
Practical implications – Even though nonusers have different reasons not to shop online, the key pattern that emerges is the value of tangibility for these individuals, which is a barrier present in all three clusters. This suggests that current marketing strategies and advertisements are ineffective to reach these consumers. Vendors should therefore try different approaches.
Originality/value – The findings contribute to the information systems (IS) literature by bringing a new perspective to the understanding of e-commerce rejection in addition to having managerial implications that involve strategies to attract potential users based on their specificities.

Keywords Consumer behavior, Online shopping, Non-shopper segmentation, Electronic market, Retailing, Shopping preference
Paper type Research paper

1. Introduction

E-commerce giants such as Amazon, Alibaba, MercadoLivre and eBay have proven how widespread the online market is. The online market is an option for retailers to expand their

The authors thank J. Figueiredo (ESPM, Escola Superior de Propaganda e Marketing) for comments on the analysis and E. Diniz for initial inspiration. They also thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for the doctoral scholarship awarded to G.N.S., and Getulio Vargas Foundation's São Paulo School of Business Administration (FGV EAESP) for resource support.

Revista de Gestão
Emerald Publishing Limited
e-ISSN: 2177-8736
p-ISSN: 1809-2276
DOI 10.1108/REGE-06-2022-0087

market and increase their profitability (G. Li, Zhang, Chiu, Liu, & Sethi, 2019), given the increasing number of online shoppers. Statista (2021b) reported that over two billion people worldwide purchased goods and services online in 2020, representing a 6.7% increase over the previous year. In Brazil, the number of e-commerce users increased by 7% between 2019 and 2021, reaching an estimated total of 68.3 million users (Regional Center for Studies on the Development of the Information Society [Cetic.br], 2022). As a result of this relevance, considerable research effort has been dedicated to understanding the profile and use patterns of online customers (H. Li, Kuo, & Russell, 1999; Pawełoszek & Korczak, 2016; Spena, D'Auria, & Bifurco, 2021; Zhou, Wei, & Xu, 2021).

The e-commerce literature focuses mostly on the technology acceptance model (TAM) (Chang, Cheung, & Lai, 2005; Iglesias-Pradas, Pascual-Miguel, Hernández-García, & Chaparro-Peláez, 2013) and concentrates on factors that affect people's intention to use e-commerce (Hsu, Chang, & Chuang, 2015; Moorthy *et al.*, 2017), their motivation for using it (Ganesh, Reynolds, Luckett, & Pomirleanu, 2010), trust or lack thereof (Bach, Silva, Souza, Kudlawicz-Franco, & Veiga, 2020; Hsu *et al.*, 2015; Maia, Lunardi, Longaray, & Munhoz, 2018), perceived risks associated with e-commerce (Bach *et al.*, 2020; Y. Li, Li, Zhang, Zhang, & Gong, 2020) or perceived benefits (Lestari, 2019). Studies have overemphasized the segmentation of e-commerce users at the expense of better understanding the segmentation of e-commerce nonusers. With the exceptions of Swinyard and Smith (2003), Anckar (2003) and Iglesias-Pradas *et al.* (2013), few studies have attempted to segment e-commerce nonusers. Scholars have rarely been concerned with classifying them based on their reasons not to shop online. This is the gap this paper attempts to address.

The failure to pay due attention to the complaints of e-commerce nonusers leads to the loss of many potential customers, especially in Brazil, which is the largest e-commerce market in Latin America with a 31.2% market share and has practically doubled its annual revenue since 2019 (Statista, 2021a). According to the Brazilian Internet Steering Committee (CGI.br, 2019), 84 million individuals in Brazil access the Internet, but do not shop online. This number represents 66% of Internet users. Their reasons include a preference for shopping in person (85%), concerns regarding personal data privacy or product/service quality (63%) and a lack of Internet skills (30%) (CGI.br, 2019).

Given this disproportion in addition to the scarce literature on e-commerce nonusers, this study advances understanding of the topic by addressing the following research questions: Do e-commerce nonusers all have the same reasons not to purchase online, or are there dissimilar behavior patterns that might lead nonusers to cluster in different groups?

To answer these questions, we used unsupervised machine learning on the answers to a nationwide survey and identified three segments of e-commerce nonusers. These segments highlight similar behavior patterns that lead nonusers to reject e-commerce. The findings contribute to the information systems (IS) literature by bringing a new perspective to the understanding of e-commerce rejection in addition to having managerial implications that involve strategies to attract potential users based on their specificities.

2. Literature review

The literature has identified several barriers to and drivers of the adoption and use of technology-dependent systems such as e-commerce, m-commerce and mobile banking. These include risk and trust (Faqih, 2016; Laumer & Eckhardt, 2012; Pavlou, 2003), value (Hsu *et al.*, 2015; Laukkanen, 2016), previous experience (Hernandez, Jimenez, & Martín, 2009), perceived cost (Moorthy *et al.*, 2017), perceived usefulness (Faqih, 2016; Lestari, 2019) and a lack of Internet skills (Iglesias-Pradas *et al.*, 2013; Scheerder, van Deursen, & van Dijk, 2017; van Deursen, Courtois, & van Dijk, 2014). From a list of twelve inhibitors of online purchasing, Anckar (2003) found four main barriers to e-commerce use: shopping limitations, cost, financial risks and search problems.

However, in the e-commerce literature, research generally addresses these barriers and drivers either from the perspective of e-commerce users (Bach *et al.*, 2020; Lestari, 2019; Y. Li *et al.*, 2020) or by comparing online shoppers and non-shoppers (Faqih, 2016; Sohail, 2014; Swinyard & Smith, 2004). Non-shoppers are seldom addressed alone. The drawback of this general approach is that studies neglect the underlying principles of e-commerce nonusers' behavior (Hernández-García, Iglesias-Pradas, Chaparro-Peláez, & Pascual-Miguel, 2011).

Scholars have argued that e-commerce nonusers are not a homogenous group (Swinyard & Smith, 2003, 2004). Some nonusers may be computer illiterate, while others are relatively digitally skilled but nevertheless do not trust e-commerce. Swinyard and Smith (2003) identified four groups of online non-shoppers: "fearful browsers," "shopping avoiders," 'technology muddlers" and "fun seekers." Their characteristics range from distrusting e-commerce to preferring to physically see the product to being computer illiterate.

Iglesias-Pradas *et al.* (2013) conducted another study and segmented e-commerce nonusers based on their reasons (barriers) not to shop online. The authors identified four types of online non-shoppers: "skeptical/distrustful," "infrastructure-conditioned," "product-conditioned" and "others." They also classified non-shoppers based on six drivers that might engage them in e-commerce: "risk-avoiders," "needers," "analog-world shoppers," "e-shopping ignorant," "hesitant" and "others." Their typology reaffirmed characteristics previously found in the literature and extended them by introducing new aspects that further distinguish e-commerce nonusers.

Studies differ in terms of the type of data analysis. In segmenting non-shoppers, Anckar (2003) and Swinyard and Smith (2003) employed factor analysis and obtained classifications based on data variance/covariance. In contrast, Iglesias-Pradas *et al.* (2013) used latent class analysis and benefited from the technique's capability to identify underlying categories and to group observations based on similar responses.

Curiously, we found no studies that have used unsupervised machine learning such as cluster analysis to **segment e-commerce nonusers**, even though the technique is a classical tool for classifying observations based on their similarities and has been successfully employed to segment e-commerce **users** (Pawełoszek & Korczak, 2016; Zhou *et al.*, 2021). Moreover, research that characterizes e-commerce nonusers based on their motives for not shopping online has decreased considerably since Iglesias-Pradas *et al.*'s (2013) study, with the aforementioned general approach prevailing. An example is Faqih's (2016) study, which, despite addressing online non-shoppers, concentrates heavily on adopting rather than rejecting online purchasing.

Finally, customer profiles change over time, as does technology; consequently, new influencing factors may emerge. Some may be accentuated, while others may become irrelevant. Thus, it is necessary to update e-commerce nonuser segmentation based on more recent surveys and by using other quantitative methods to verify what has changed over the years. Therefore, we address the segmentation of e-commerce nonusers by employing cluster analysis of a list of reasons reported by Internet users that prevent them from purchasing online.

3. Materials and methods

All statistical analyses were performed in RStudio 1.3.959. The data used in this study are freely available on the Cetic.br (2019) website and are used aggregately.

3.1 Sample selection and description

This study used the Information and Communication Technology (ICT) Households 2018 microdata from the Cetic.br (2019) website, a CGI.br associated department. The ICT

Households survey is a nationwide survey that has been conducted since 2005 with the objective of measuring the availability, possession and use of ICT by the Brazilian population aged ten years and older. Data collection occurs through structured questionnaires composed of closed questions with defined answers that are administered in face-to-face interviews with the respondents (CGI.br, 2019). The sample is representative of the country and consisted of 20,544 respondents in the 2018 edition. For this paper, we used the subsample of respondents who reported their motives for not buying goods or services online (9,522 respondents, 46.3% of the total).

In the ICT Households 2018 e-commerce module, indicator H2 contains the reports of respondents who bought online and those who did not. We used the respondents who reported not buying online. A negligible number of respondents (<1%) did not answer question H2 and were considered online non-shoppers. Exploratory analyses identified 457 outliers, who were removed to enhance the formation of clusters. Thus, the final sample contained 9,065 respondents. Of those respondents, 42.8% were male. The average age was 35.9 ± 17.1 (mean ± SD). Levels of education were divided into illiterate/preschool (4%), elementary education (43%), secondary education (43.5%) and tertiary education (9.5%). Family income was distributed among no income (24.6%), up to the minimum wage (MW, 43.7%), more than the MW and up to twice the MW (20.8%), more than twice the MW and up to three times the MW (6.3%), more than three times the MW and up to five times the MW (3.3%), more than five times the MW and up to ten times the MW (1%), and more than ten times the MW (0.3%). With regard to social class, 14.5% were upper class (classes A and B), 50.5% were middle class (class C) and 35% were lower class (classes D and E). Finally, 91% of respondents lived in urban areas.

3.2 Exploratory analysis and cluster formation

Initially, nine dichotomous variables (Table 1) addressed reasons why respondents decided not to purchase online. A preference for shopping in person and concerns regarding personal data privacy or product/service quality were prominent. In other words, tangibility, risk and lack of trust were the most frequent factors people indicated as motives for not shopping online.

Based on the literature, we created **four indicators** that grouped the variables into broader dimensions, which are detailed in subsections 3.2.1 through 3.2.4. The variables "lack of need," "lack of interest" and "preference for shopping in person" were grouped and labeled the "disinterest" dimension. The variable "lack of Internet skills" was not grouped and was

Table 1. Variables in the analysis

Variable code	Reasons why the respondent did not buy or order products and services on the Internet	Quantity	%
H3_A1	Lack of need	4,377	48
H3_B1	Lack of interest	5,359	59
H3_C1	Preference for shopping in person and seeing the product	7,679	85
H3_D1	Lack of Internet skills	2,742	30
H3_E1	Delivery takes too long or it is difficult to receive products at home	4,394	48
H3_F1	Concerns about security and privacy, or about providing personal information	5,738	63
H3_G1	Impossibility of making online payments	3,585	40
H3_H1	Lack of trust in the product that will be received	5,933	65
H3_I1	Impossibility of making complaints or returning the product	4,811	53

Note(s): N = 9,065 in every row. The quantity and percentage of "yes" for each variable are reported
Source(s): Table by authors

labeled the "inability" dimension. The variables "delivery takes too long," "impossibility of making online payments" and "impossibility of making complaints" were grouped and labeled the "operational difficulty" dimension. Finally, the variables "concerns about security" and "lack of trust" were grouped and labeled the "distrust" dimension. Table 2 summarizes the indicators. Each dimension is explained next.

3.2.1 Indicator 1: disinterest dimension. Consumer behavior and attitudes toward a given product or service are complex matters. Several factors have been indicated as influencing people's behavior, such as external conditions (Guagnano, Stern, & Dietz, 1995) or values, beliefs and norms (Zepeda & Deal, 2009). In the IS field, Lestari (2019) confirmed that perceived usefulness affects people's intention to adopt e-commerce. Hernández-García *et al.* (2011) found that perceived compatibility, that is, whether an individual believes that the technology is compatible with him/her, is the strongest factor that influences non-shoppers' attitudes toward e-commerce. Furthermore, Laukkanen (2016) focused on consumer rejection of a given technology; Faqih (2016) investigated the factors that slow Jordanian Internet users' adoption of online purchases; and Mainardes, Souza, and Correia (2020) recently investigated why people show disinterest in e-commerce.

From these studies and others, we know that previous experiences (Hernandez *et al.*, 2009), tangibility (Iglesias-Pradas *et al.*, 2013; Liu & Wei, 2003; Swinyard & Smith, 2003) and value (Hsu *et al.*, 2015; Laukkanen, 2016) account for why some people are not interested in purchasing online. Put simply, they do not see value in e-commerce. As we have seen, three of the CGI.br (2019) variables measured these factors (H3_A1, B1 and C1). Given their similarity, we grouped them into the same indicator, which was labeled "disinterest" and consisted of the sum of the variables (range: 0–3).

3.2.2 Indicator 2: inability dimension. The variable H3_D1 measured whether a lack of Internet skills is a motive for not purchasing online. Although one's inability to use ICT tools or access the Internet may lead to operational difficulties when shopping online, this does not necessarily mean it is an operational difficulty; rather, it reflects one's digital skills. Indeed,

Code	Item	Description	Indicator	Theoretical support
H3_A1	Lack of need	Lack of interest, need, and/or preference	Disinterest (0–3)	Zepeda and Deal (2009), Iglesias-Pradas *et al.* (2013), Mainardes *et al.* (2020)
H3_B1	Lack of interest			
H3_C1	Preference for shopping in person and seeing the product			
H3_D1	Lack of Internet skills	Lack of Internet skills (original preserved)	Inability (0–1)	Pavlou (2003), van Deursen *et al.* (2014), Scheerder *et al.* (2017)
H3_E1	Delivery takes too long or it is difficult to receive products at home	Troubles with receiving, paying for, and/or returning the product	Operational difficulty (0–3)	Laukkanen (2016), Zhu and Chen (2013), Nery-da-Silva *et al.* (in press)
H3_G1	Impossibility of making online payments			
H3_I1	Impossibility of making complaints or returning the product			
H3_F1	Concerns about security and privacy or about providing personal information	Afraid of providing information or concerns about product/service quality	Distrust (0–2)	Liu and Wei (2003), Pavlou (2003), Li *et al.* (2020)
H3_H1	Lack of trust in the product that will be received			

Source(s): Table by authors

Table 2. Summary of the indicators

there is a specific body of literature on digital skills alone (Araujo & Reinhard, 2018; Scheerder *et al.*, 2017; van Deursen *et al.*, 2014; van Deursen, Helsper, & Eynon, 2016), and treating digital skills as an operational limitation would be a mistake. Therefore, to avoid improper mixing, this variable represented its own dimension and was labeled "inability." Its dichotomous characteristic was retained in the calculations.

3.2.3 Indicator 3: operational difficulty dimension. The items that constituted this indicator measured difficulties concerning paying for, receiving or complaining about products/services (Table 2). These factors are related to operational issues; for example, given logistical or bureaucratic requirements or insecurity, an individual would prefer to avoid purchasing online rather than risk doing so and subsequently regretting it. Previous experiences (Hernandez *et al.*, 2009), social influence (Mainardes *et al.*, 2020), brand image (Laukkanen, 2016), computer literacy (Swinyard & Smith, 2003) and area of residence (Nery-da-Silva, Barbosa, & Figueiredo, in press; Zhu & Chen, 2013) are examples of restrictions that lead people to avoid shopping online or adopt some technologies. Thus, the variables H3_E1, G1 and I1 were grouped to form the indicator "operational difficulty," which consisted of the sum of the variables (range: 0–3).

3.2.4 Indicator 4: distrust dimension. Pavlou (2003) extensively addressed trust and risk in e-commerce. He proposed the e-commerce acceptance model by integrating trust and risk into the TAM. According to his findings, trust positively affects people's intentions to transact online and actual online transactions and negatively affects perceived risk. In turn, the latter negatively affects users' intentions to transact online. Other studies have also found that perceived risk negatively affects people's intentions to adopt e-commerce (Bach *et al.*, 2020; Y. Li *et al.*, 2020; Liu & Wei, 2003). These studies confirm that people are wary about providing their data to web retailers, which is confirmed by the CGI.br's surveys. This wariness may generate a lack of trust in websites or the system itself. Given that variables H3_F1 and H3_H1 address this subject matter, they were grouped to form an indicator that was labeled "distrust" and consisted of the sum of the variables (range: 0–2).

3.3 Cluster analysis

Cluster analysis was conducted in two stages: (i) hierarchical clustering analysis (HCA) and (ii) partitioning around medoids (PAM) algorithm. For the HCA technique, we employed Gower's (1971) association metric with average linkage algorithm. We chose this method rather than Ward's because the latter tends to create clusters with approximately the same number of observations (Hair, Black, Babin, & Anderson, 2019), and we were interested in identifying variations in cluster sizes. The cophenetic correlation coefficient (Sokal & Rohlf, 1962) also influenced our choice of the average linkage method. The cut tree was determined by studying the dendrogram and heights and considering silhouette method estimates (Kaufman & Rousseeuw, 1990).

To optimize the results, we followed Hair *et al.*'s (2019) suggestion of combining hierarchical and nonhierarchical cluster analyses. The PAM algorithm was employed in the nonhierarchical clustering analysis because this algorithm is less sensitive to noise and outliers than other methods (Kassambara, 2017; Kaufman & Rousseeuw, 1990).

3.4 Cluster validation statistics

To assess cluster validation, we verified the cophenetic correlation coefficient; within-cluster sum of squares (WSS); average within, between and silhouette widths; Dunn index; and Pearson-Gamma (Halkidi, Batistakis, & Vazirgiannis, 2001). Additionally, we employed one-way analysis of variance (ANOVA) with post hoc analyses and chi-square tests to assess the discriminating power of the clusters.

4. Results

The HCA was performed with 9,065 respondents. By examining the dendrogram, we noted three clear clusters, which were also confirmed by the silhouette method. Cluster 1 (C1) had 2,742 members (30%), cluster 2 (C2) had 2,623 members (29%) and cluster 3 (C3) had 3,700 members (41%). Cophenetic correlation suggested that clustering showed strong fidelity to the original data ($c = 0.76$). HCA was conducted mainly to guide us through PAM analysis.

The PAM analysis results were highly consistent with the HCA results (see Figure 2A). The coincidence rates were 98%, 76% and 96% for C1, C2 and C3, respectively. The average distance within clusters was 0.24, suggesting relatively satisfactory cluster compactness, and the average distance between clusters was 0.55, which was more than twice that of the within-cluster distance and suggested that there were relatively large distances between the clusters. There was low variability in the observations within clusters, with WSS = 344.08, which is a very small number considering the sample size. The cluster average silhouette widths also indicated relatively good clustering (0.57 for C1, 0.33 for C2, and 0.55 for C3), with an overall average of 0.47. Finally, the Pearson-Gamma indicated a strong association among cluster members ($\Gamma = 0.67$), but the Dunn index indicated that the clusters were not as compact and well separated (Dunn index = 0.08) as they could be.

The discriminating power of the clusters was verified for each dimension (see Table 3 for the central tendency measures and variability). The differences in the variances of the disinterest frequencies were statistically significant between clusters ($H(2) = 1434.75$, $p < 0.001$, Kruskal–Wallis rank-sum test; $\alpha = 0.05$ for Dunn's multiple comparison). Similarly, the differences in the variances of the operational difficulty frequencies were statistically significant between clusters ($H(2) = 4232.29$, $p < 0.001$, Kruskal–Wallis rank-sum test; $\alpha = 0.05$ for Dunn's multiple comparison). The variance of the frequencies in the distrust dimension was significantly different between C1 and C3 ($H(2) = 5984.03$, $p < 0.001$, Kruskal–Wallis rank-sum test and $\alpha = 0.05$ for Dunn's multiple comparison) and between C2 and C3 ($p < 0.001$) but not between C1 and C2 ($p > 0.05$). Finally, differences in frequencies in the inability dimension were also significantly different between clusters ($\chi^2(2, N = 9{,}065) = 9{,}065$, $p < 0.001$). Altogether, the results suggest the clusters identified the members well, confirming the existence of heterogeneous groups of nonusers in the Brazilian e-commerce context.

4.1 Overview of the clusters

Table 3 shows the cluster means and medoids for each indicator. C1 is the cluster in which nearly every reason on the list accounts for its members not purchasing goods and services on the Internet. In contrast, C3 is represented by respondents who scored zero in all but one dimension, suggesting that one factor in the disinterest dimension is the main reason why the members of this cluster reject e-commerce.

	Cluster 1		Cluster 2		Cluster 3	
Dimension	Mean (SD)	Medoid	Mean (SD)	Medoid	Mean (SD)	Medoid
Disinterest	2.29 (0.82)	*3*	2.05 (0.80)	2	1.40 (0.86)	*1*
Inability	–	*1*	–	0	–	0
Operational difficulties	2.12 (0.99)	2	1.77 (0.92)	2	0.28 (0.52)	0
Distrust	1.75 (0.43)	2	1.74 (0.47)	2	0.27 (0.45)	0
Size (%)	2742 (30%)		3507 (39%)		2816 (31%)	

Note(s): $N = 9{,}065$
Source(s): Table by authors

Table 3. Cluster means and medoids

Regarding frequencies, C1 is mostly characterized by respondents who affirm that all variables in the disinterest dimension lead them to reject e-commerce (Table 3 and Figure 1A), and it is the only cluster that indicates inability as a reason not to buy online (Table 3 and Figure 1B). It also scores the highest in both the operational difficulty dimension (Figure 1C) and the distrust dimension (Figure 1D). C2 shares many similarities with C1 in three out of the four dimensions (Figure 1A, C, and D; also see the medoids in Table 3). It tends to represent two motives in both the disinterest dimension (Figure 1A) and the distrust dimension (Figure 1D) and oscillates between one and two operational difficulties as motives for rejecting e-commerce (Figure 1C).

C3 is considerably different from the other clusters. It can be easily noticed in Figure 1A and D that it stands out in scoring at the bottom of the scales in all but one dimension.

We also explored the descriptive characteristics of each cluster after they were formed. There is a balanced distribution of members across the clusters, with C2 being the largest (Figure 2A). C1 differs from the others in terms of age. It consists of older people (average age = 43.3; Figure 2B), whereas the others present an average age of approximately 30 (Figure 2B). C3 has a modal age of eleven, which is likely associated with the cluster characteristics.

Figure 2C depicts the social class distribution in each cluster. We can see little participation among the upper class (A-B) in C1 and some equilibrium between the other classes in the same cluster. The upper class is represented similarly in C2 and C3 (~50%), and the middle class (class C) is distributed equally across the clusters. The lower class (D-E) tends to belong to C1 rather than to C2 and C3. Regarding the distribution of areas of residence, no relevant differences among the clusters were found (Figure 2D). The proportion of urban residents across clusters was 91 $\pm$ 0.02%.

However, socioeconomic factors varied more across clusters. The charts in Figure 2C, E, and F show that socioeconomic attributes may be associated with the cluster to which one belongs. Based on the medoid values in Table 3, C1 is the cluster in which almost every reason comprises members' decisions not to purchase online. In Figure 2E (left), C1 consists more of illiterate members (7%) than the others do, and most of the members have elementary education (54%); in addition, it is the only cluster affected by inability issues. This result is consistent with previous studies suggesting that a lack of digital skills is more expected from less educated people (Araujo & Reinhard, 2018; CGI.br, 2019).

In C2, we see greater participation among people with a secondary or tertiary education (Figure 2E, center). Compared to C1, the proportion of elementary-educated individuals falls by 20%, whereas that of secondary-educated individuals rises by 18%, which is practically an inversion of the proportions from C1 to C2. Additionally, tertiary-educated people are more common in C2 than in C1.

C3 has a balanced distribution of proportions between elementary- and secondary-educated people (43% and 41%, respectively; Figure 2E, right). Tertiary-educated

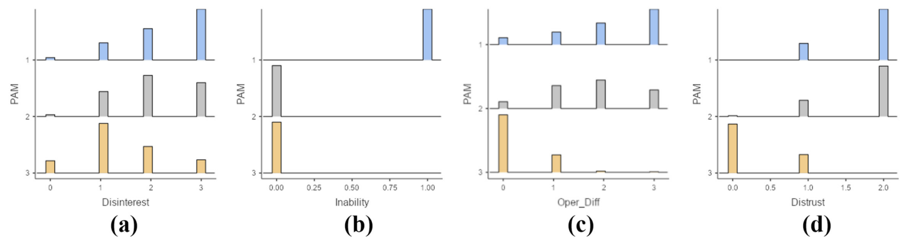
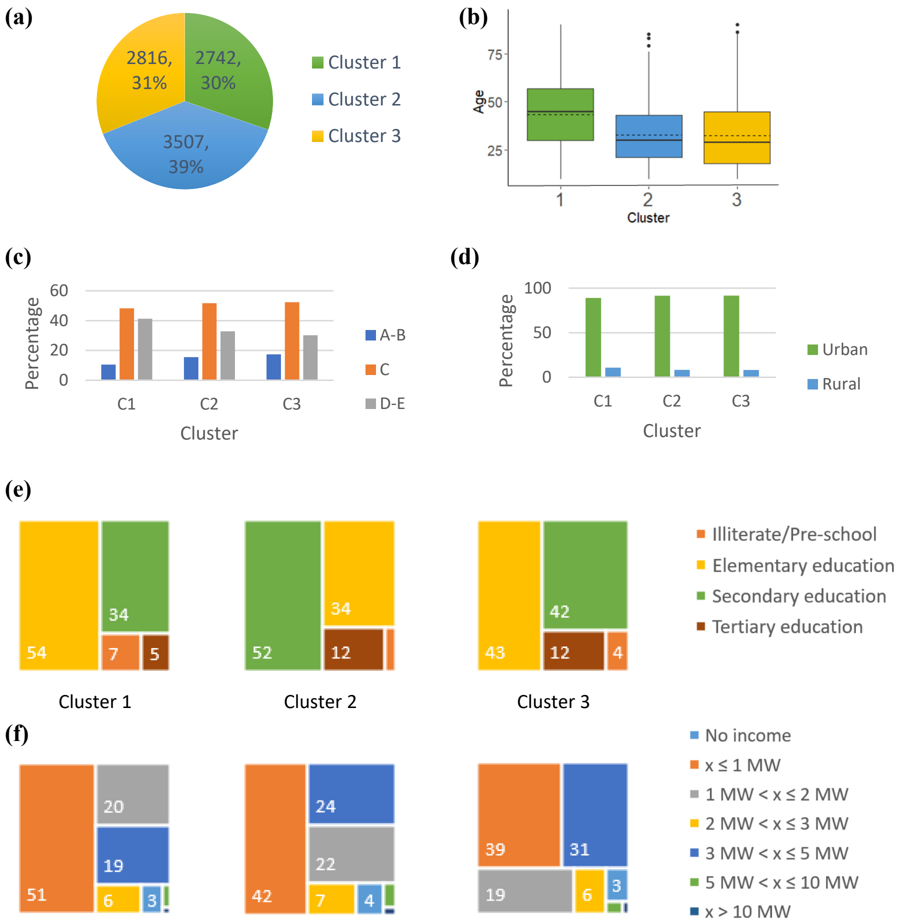


Figure 1. Visual distribution of frequencies of the clusters in each indicator. Bar chart of frequencies of each cluster in (A) disinterest dimension, (B) inability dimension, (C) operational difficulty dimension, (D) distrust dimension

Source(s): Figure by authors



Source(s): Figure by authors

Figure 2. An overview of the clusters and their description. (A) Pie chart showing the distribution of the sample between the clusters. (B) Boxplots comparing the age in each cluster. For C1, mean = 43.3 ± 17.1, mode = 48, median = 45; for C2, mean = 32.9 ± 15, mode = 20, median = 30; and for C3, mean = 32.4 ± 17.1, mode = 11, median = 29. (C) Social class distribution in each cluster. (D) Distribution of the area of residence in each cluster. (E) Proportion of the level of education (given in percentage) in C1 (left), C2 (center) and C3 (right). (F) Proportion of family income (given in percentage) in C1 (left), C2 (center) and C3 (right)

individuals constitute 12% of the total, just as in C2. Illiterate people are slightly more present in C3 than in C2, perhaps because of the large presence of younger people in C3. Therefore, most illiterate individuals, when considering their social position and not their age, are in C1.

Taken together, these results suggest that education is a factor associated with the decision to use e-commerce. It is reasonable to assume that the lower an individual's educational level is, the more he or she tends to overestimate the process of purchasing online, leading him/her to exaggeratedly affirm that each reason on a given list accounts for why he/she rejects e-commerce rather than benefiting from it.

Finally, in Figure 2F, we plot the distribution of family income in each cluster. Irrespective of the proportion, income from zero up to twice the MW is consistently the three most frequent income strata in all clusters. However, there is a slightly higher presence of high-income earners in C2 (Figure 2F, center), suggesting that high-income earners tend to belong more to C2. C3 has a higher proportion of incomeless members (31%; Figure 2F, right) than C1 and C2, which is most likely explained by the high presence of young members in that cluster (see Figure 2B).

4.2 Understanding the clusters: main features and names

C1 had 2,742 (30%) members, consisted of 61% female respondents and had an average age of 43.3 ± 17.1. The members of this cluster scored the highest in all the dimensions, standing out in the disinterest and inability dimensions. They are confused about purchasing online since all the reasons seem to affect them. Furthermore, they are resistant to technology and have more trouble with it. Considering all these features, C1 was named **reluctant ones**.

C2 had 3,507 (39%) members, consisted of 58% female respondents and had an average age of 32.9 ± 15. According to the scores in each dimension, the members of this cluster scored either near the average or above it in every dimension except inability. In other words, they did not stand out in any dimension but scored in all of them, suggesting that e-commerce does not work for them. Thus, C2 was named **disbelievers** because even though its members are not restricted by a lack of digital skills, they are not interested in e-commerce, nor do they have the desire to give up tangibility. They do not want to wait for a product to arrive, and they are not willing to deal with problems if something goes wrong. They do not want to provide their data, nor do they trust that the product will be delivered. In summary, they do not believe e-commerce works.

Finally, C3 had 2,816 (31%) members, consisted of 53% female respondents and had an average age of 32.4 ± 17.1. Its main characteristic, as shown in Table 3, is that the members of this cluster are not interested in purchasing online. They scored the highest on the variable "preference for shopping in person" (46%), which means that tangibility is, by far, the most critical factor for them. Moreover, this group values shop tangibility and human interactions, meaning they prefer to speak to an actual person over chatting with a bot, even to hire a service, which is not tangible at all. For these reasons, C3 was named **doubting Thomas** because they have to see it to believe it.

5. Discussion and conclusion

Previous attempts to segment e-commerce nonusers found four types of online non-shoppers (Iglesias-Pradas *et al.*, 2013; Swinyard & Smith, 2003) or four factors that constitute barriers to e-commerce access (Anckar, 2003). In our study, three clusters of nonusers were identified from four dimensions of reasons not to shop online. **Reluctant ones** are the most intriguing group. Several demographic factors historically associated with technology adoption (Li *et al.*, 1999; Venkatesh, Thong, & Xu, 2012; Zhu & Chen, 2013) are present in the cluster. The members tend to be less educated, lower class and older than the members of the other clusters. Additionally, they are impacted by all dimensions. Compared with Swinyard and Smith's classification, **reluctant ones** are a mix of shopping avoiders and technology muddlers.

The variety of characteristics in this group poses challenges regarding how to engage its members in online shopping. All of these factors have been identified by previous studies (Laukkanen, 2016; Venkatesh *et al.*, 2012; Zhu & Chen, 2013), so they are not novel; nevertheless, they continue to affect people in terms of technology adoption. To address the dimensions considering these demographics, we recommend that online merchants should focus their actions on the barriers that are common to most of the clusters, that is, distrust and tangibility.

Likewise, the existence of the **disbelievers** group reinforces the need for more accurate marketing strategies, better customer care, better delivery systems and more attention to customer experience and engagement. Our findings emphasize that e-commerce nonusers are complex and that simple nonuser labeling is too narrow to represent their diversity. They do not buy online because different factors may or may not influence their behavior. These particularities must be considered to design effective, customized strategies.

The **disbelievers** cluster roughly brings together three types of non-shoppers from Iglesias-Pradas *et al.*'s (2013) study, namely skeptical/distrustful, risk-avoiders and hesitant. In both cases, despite the resistance identified, there is still room to engage the members in online shopping due to drivers that can be used as motivations for them to make online purchases, such as helping them know the seller better, facilitating the perception of security in online transactions and reducing their skepticism about product tangibility.

Finally, we also found the **doubting Thomas** cluster. A curious phenomenon occurs in this cluster: it has proportionally more members from the upper and lower classes than the other clusters (Figure 2C), has more illiterate individuals than the **disbelievers** cluster (Figure 2E, right), has a modal age of eleven but an average age of 32.4 and is mostly affected by only one of the dimensions of reasons not to shop online (Table 3). The fact that its modal age is eleven explains the higher proportion of incomeless members compared to the other clusters (Figure 2F) and why its members do not score in almost any dimension investigated. Because they are younger (from Generation Y on), they are less afraid to provide personal information (Gewald *et al.*, 2017) and are usually more digitally skilled (Lissitsa & Kol, 2016). Nonetheless, the members of the **doubting Thomas** cluster value tangibility.

Previous studies have documented that individuals may resist, not adopt (Hernandez *et al.*, 2009; Laumer & Eckhardt, 2012) or not be interested in technology (Mainardes *et al.*, 2020) due to computer-literacy obstacles, perceived ability or lack of trust (Faqih, 2016; Liu & Wei, 2003; Pavlou, 2003; Swinyard & Smith, 2003). Our findings are aligned with these studies as they highlight the role these factors play in leading many individuals toward such attitudes.

In conclusion, our findings identify three clusters of e-commerce nonusers. These nonusers have different reasons not to shop online, but the key pattern that emerges is the value of tangibility for these individuals, which is a barrier present in all three clusters. This suggests that current marketing strategies and advertisements are ineffective to reach these consumers. A different approach should be used that focuses on the particularities present in each cluster.

Vendors should improve the virtual experience by investing in augmented reality. They could also work with showrooming and webrooming as a first step to engage these consumers and establish easy product return policies for online purchases. To reduce distrust and address computer-literacy obstacles, governments can provide basic digital literacy and promote the benefits of online shopping to help people ensure that online transactions are safe.

Our study also provides methodological contributions by bringing to light the utility of unsupervised machine learning in segmenting non-shoppers, particularly by revealing the underlying patterns of reasons not to shop online.

Despite the insights obtained, one limitation is the cross-sectional design, which restricts the understanding of the phenomenon. We recommend that future studies replicate our analysis with longitudinal data and check for changes in behavior patterns over time. We also encourage researchers to undertake analyses to determine what other factors have the strongest effects on non-shoppers, particularly by examining what may have changed since the outbreak of COVID-19. Perhaps people have become more susceptible to engaging in the online market, and some factors may no longer be relevant.

References

Anckar, B. (2003). Drivers and inhibitors to E-commerce adoption: Exploring the rationality of consumer behavior in the electronic marketplace. In *ECIS 2003 Proceedings*, Paper 24.

Araujo, M. H., & Reinhard, N. (2018). Caracterizando os usuários de Internet no Brasil: uma análise a partir das habilidades digitais. In *Twenty-Fourth Americas Conference on Information Systems*, New Orleans (1–10), August.

Bach, T. M., Silva, W. V., Souza, A. M., Kudlawicz-Franco, C., & Veiga, C.P. (2020). Online customer behavior: Perceptions regarding the types of risks incurred through online purchases. *Palgrave Communications*, *6*(1). doi: 10.1057/s41599-020-0389-4.

Cetic.br (2019). [Microdados] TIC Domicílios - 2019 – Indivíduos. Available from: https://www.cetic.br/pt/pesquisa/domicilios/microdados/ (accessed 12 August 2020).

Cetic.br (2022). TIC Domicílios 2021: Lançamento dos resultados. Available from: https://cetic.br/media/analises/tic_domicilios_2021_coletiva_imprensa.pdf (accessed 16 October 2022).

CGI.br (2019). *Survey on the use of information and communication technologies in brazilian households: ICT households 2018 [ebook] (Bilingual)*. São Paulo: CGI.br.

Chang, M. K., Cheung, W., & Lai, V. S. (2005). Literature derived reference models for the adoption of online shopping. *Information & Management*, *42*(4), 543–559. doi: 10.1016/j.im.2004.02.006.

Faqih, K. M. S. (2016). An empirical analysis of factors predicting the behavioral intention to adopt Internet shopping technology among non-shoppers in a developing country context: Does gender matter? *Journal of Retailing and Consumer Services*, *30*, 140–164. doi: 10.1016/j.jretconser.2016.01.016.

Ganesh, J., Reynolds, K. E., Luckett, M., & Pomirleanu, N. (2010). Online shopper motivations, and e-store attributes: An examination of online patronage behavior and shopper typologies. *Journal of Retailing*, *86*(1), 106–115. doi: 10.1016/j.jretai.2010.01.003.

Gewald, H., Wang, X., Weeger, A., Raisinghani, M. S., Grant, G., Sanchez, O. P., & Pittayachawan, S. (2017). Millennials' attitudes toward IT consumerization in the workplace. *Communications of the ACM*, *60*(10), 62–69. doi: 10.1145/3132745.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871.

Guagnano, G. A., Stern, P. C., & Dietz, T. (1995). Influences on attitude-behavior relationships: A natural experiment with curbside recycling. *Environment and Behavior*, *27*(5), 699–718. doi: 10.1177/0013916595275005.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Andover: Cengage Learning.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2-3), 107–145. doi: 10.1023/A:1012801612483.

Hernández-García, Á., Iglesias-Pradas, S., Chaparro-Peláez, J., & Pascual-Miguel, F. (2011). Exploring the attitudes and intentions of non-shoppers in the acceptance of e-commerce. *Journal of Universal Computer Science*, *17*(9), 1314–1328.

Hernandez, B., Jimenez, J., & Martín, M. J. (2009). Adoption vs acceptance of e-commerce: Two different decisions. *European Journal of Marketing*, *43*(9), 1232–1245. doi: 10.1108/03090560910976465.

Hsu, M. H., Chang, C. M., & Chuang, L. W. (2015). Understanding the determinants of online repeat purchase intention and moderating role of habit: The case of online group-buying in Taiwan. *International Journal of Information Management*, *35*(1), 45–56. doi: 10.1016/j.ijinfomgt.2014.09.002.

Iglesias-Pradas, S., Pascual-Miguel, F., Hernández-García, Á., & Chaparro-Peláez, J. (2013). Barriers and drivers for non-shoppers in B2C e-commerce: A latent class exploratory analysis. *Computers in Human Behavior*, *29*(2), 314–322. doi: 10.1016/j.chb.2012.01.024.

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (1st ed.). STHDA, available at: https://www.datanovia.com/en/product/practical-guide-to-cluster-analysis-in-r/.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Laukkanen, T. (2016). Consumer adoption versus rejection decisions in seemingly similar service innovations: The case of the Internet and mobile banking. *Journal of Business Research*, *69*(7), 2432–2439. doi: 10.1016/j.jbusres.2016.01.013.

Laumer, S., & Eckhardt, A. (2012). Why do people reject technologies: A review of user resistance theories. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), *Information Systems Theory: Explaining and Predicting Our Digital Society* (Vol. 1, pp. 63–86). New York: Springer. doi: 10.1007/978-1-4419-6108-2_4.

Lestari, D. (2019). Measuring e-commerce adoption behaviour among gen-Z in Jakarta, Indonesia. *Economic Analysis and Policy*, *64*, 103–115. doi: 10.1016/j.eap.2019.08.004.

Li, H., Kuo, C., & Russell, M. G. (1999). The impact of perceived channel utilities, shopping orientations, and demographics on the consumer's online buying behavior. *Journal of Computer-Mediated Communication*, *5*(2). doi: 10.1111/j.1083-6101.1999.tb00336.x.

Li, Y., Li, X., Zhang, Z., Zhang, G., & Gong, M. (2020). Understanding consumers online furniture purchase behavior: An updated UTAUT perspective. *Journal of Forest Economics*, *35*(4), 267–303. doi: 10.1561/112.00000516.

Li, G., Zhang, X., Chiu, S. M., Liu, M., & Sethi, S. P. (2019). Online market entry and channel sharing strategy with direct selling diseconomies in the sharing economy era. *International Journal of Production Economics*, *218*, 135–147. doi: 10.1016/j.ijpe.2019.05.006.

Lissitsa, S., & Kol, O. (2016). Generation X vs. Generation Y - a decade of online shopping. *Journal of Retailing and Consumer Services*, *31*, 304–312. doi: 10.1016/j.jretconser.2016.04.015.

Liu, X., & Wei, K. K. (2003). An empirical study of product differences in consumers' E-commerce adoption behavior. *Electronic Commerce Research and Applications*, *2*(3), 229–239. doi: 10.1016/S1567-4223(03)00027-9.

Maia, C., Lunardi, G., Longaray, A., & Munhoz, P. (2018). Factors and characteristics that influence consumers' participation in social commerce. *Revista de Gestão*, *25*(2), 194–211. doi: 10.1108/REGE-03-2018-031.

Mainardes, E. W., Souza, I. M., & Correia, R. D. (2020). Antecedents and consequents of consumers not adopting e-commerce. *Journal of Retailing and Consumer Services*, *55*(April). doi: 10.1016/j.jretconser.2020.102138.

Moorthy, K., Suet Ling, C., Weng Fatt, Y., Mun Yee, C., Ket Yin, E. C., Sin Yee, K., & Kok Wei, L. (2017). Barriers of mobile commerce adoption intention: Perceptions of generation X in Malaysia. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(2), 37–53. doi: 10.4067/S0718-18762017000200004.

Nery-da-Silva, G., Barbosa, R. A. P., & Figueiredo, J. C. B. (in press). The impact of socio-economic and demographic factors on the individual use of e-commerce in Brazil: Evidence of the digital divide. *International Journal of Business Information Systems*, 1–17. doi: 10.1504/IJBIS.2022.10049394.

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, *7*(3), 101–134. doi: 10.1080/10864415.2003.11044275.

Pawełoszek, I., & Korczak, J. (2016). An approach to discovery of customer profiles. In A. Tjoa, L. Xu, M. Raffai, & N. Novak (Eds.), *CONFENIS 2016: Research and Practical Issues of Enterprise Information Systems* (Vol. 268, pp. 88–99). Springer. doi: 10.1007/978-3-319-49944-4_7, available at: https://link.springer.com/chapter/10.1007/978-3-319-49944-4_7#citeas.

Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. *Telematics and Informatics*, *34*(8), 1607–1624. doi: 10.1016/j.tele.2017.07.007.

Sohail, M. S. (2014). Factors Impeding online shopping: An arab world perspective. *European Online Journal of Natural and Social Sciences*, *3*(4), 972–982. Available from: www.european-science.com

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, *11*(2), 33–40. Available from: https://www.jstor.org/stable/1217208

Spena, T. R., D'Auria, A., & Bifurco, F. (2021). Customer insight and customer profiling. In T. R. Spena, & F. Bifurco (Eds.), *Digital Transformation in the Cultural Heritage Sector: Challenges to Marketing in the New Digital Era* (pp. 95–117). Cham: Springer Nature.

Statista (2021a). E-Commerce in Brazil - statistics & facts. Available from: https://www.statista.com/topics/4697/e-commerce-in-brazil/#dossierKeyfigures (accessed Statista website accessed 20 October 2021).

Statista (2021b). E-Commerce worldwide - statistics & facts. Available from: https://www.statista.com/topics/871/online-shopping/ (accessed Statista website accessed 2 September 2021).

Swinyard, W. R., & Smith, S. M. (2003). Why people (Don't) Shop Online: A Lifestyle Study of the Internet Consumer. *Psychology & Marketing*, *20*(7), 567–597. doi: 10.1002/mar.10087.

Swinyard, W. R., & Smith, S. M. (2004). Activities, Interests, And Opinions Of Online Shoppers And Non-Shoppers. *International Business & Economics Research Journal*, *3*(4), 37–48.

van Deursen, A. J. A. M., Courtois, C., & van Dijk, J. A. (2014). Internet Skills, Sources of Support, and Benefiting From Internet Use. *International Journal of Human-Computer Interaction*, *30*(4), 278–290. doi: 10.1080/10447318.2013.858458.

van Deursen, A. J. A. M., Helsper, E. J., & Eynon, R. (2016). Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society*, *19*(6), 804–823. doi: 10.1080/1369118X.2015.1078834.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly: Management Information Systems*, *36*(1), 157–178. doi: 10.2307/41410412.

Zepeda, L., & Deal, D. (2009). Organic and local food consumer behaviour: Alphabet theory. *International Journal of Consumer Studies*, *33*(6), 697–705. doi: 10.1111/j.1470-6431.2009.00814.x.

Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services*, *61*, 102588. doi: 10.1016/j.jretconser.2021.102588.

Zhu, S., & Chen, J. (2013). The digital divide in individual e-commerce utilization in China: Results from a national survey. *Information Development*, *29*(1), 69–80. doi: 10.1177/0266666912450168.

Corresponding author
Gabriel Nery-da-Silva can be contacted at: nery.gal.nery@gmail.com