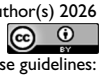
Check for updates

Research Paper

JIS

Journal of Information Science
1–22
© The Author(s) 2026

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01655515261458267
journals.sagepub.com/home/jis
Sage

A data-driven approach to increase online readers' engagement and subscribers

Elizabeth Fernandes[1], Sérgio Moro[1] and Paulo Cortez[2]

Abstract

The popularization improved the trend towards visual social media platforms, presenting novel strategic opportunities for the media industry. This study presents a data-driven framework for visual content strategies, addressing a notable research gap regarding the strategic utilization of Instagram by news organizations. Drawing on internal and external data sets gathered during the COVID-19 pandemic, the research employs a mixed-methods approach comprising longitudinal statistical analysis, computational sentiment analysis, and quantitative discussion. The findings demonstrate that narrative engagement plays an important role in involving audiences. Instagram readers on the website increased, as well as their engagement, revealing the assertiveness in content selection. The results underscore the positive effect of vivid content and illustrate the strategic advantages of integrating a data-driven decision-making framework into editorial workflows. These results assist social media managers in defining an editorial strategy to maximize reach and enhance reader conversion.

Keywords

data-driven strategy; data analysis; digital journalism; social media; data visualization; visual content

1. Introduction

In the current digital landscape, the pervasiveness of real-time reader interaction generates vast quantities of granular data for media organizations. This shift into the big data era has necessitated the adoption of sophisticated technologies, integrating data science (DS) and business intelligence (BI) capabilities directly into publishing workflows [1,2]. Consequently, the decision-making processes governing content distribution and promotion have grown increasingly multifaceted. Within this framework, data visualization and analytical modelling have become pivotal instruments for optimizing marketing strategies and editorial interventions across the reader conversion funnel [3].

In the last decade, publishers faced a digital transformation that introduced challenges and opportunities to the media industry [1,4]. As people engage with social media (SM) platforms [5–7], news organizations have had to adapt their strategies to engage followers and boost readers' engagement on their own platforms [8]. Furthermore, the increase in visual content published online has induced communication researchers to study this recent reality [9]. As argued by Araujo et al. [9], news image analysis is of interest to all sorts of communication researchers. However, the role and strategy of specific SM content is understudied [9,10].

Authors argue that audiences are sensitive to the tone and style of communication [11]. Furthermore, news organizations have been embracing the use of analytics as part of editorial decision-making in a big data environment [2,12]. Technological tools [13], organizational structures, and cultural components were necessary parts to build a prodata

[1]ISTAR, ISCTE – Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal
[2]ALGORITMI Research Centre, University of Minho, Guimarães, Portugal

Corresponding author:
Elizabeth Fernandes, ISTAR, ISCTE – Instituto Universitário de Lisboa (ISCTE-IUL), Avenida das Forças Armadas, Edifício II, D615, 1649-026 Lisboa, Portugal.
Email: elisabeth.ferna@gmail.com

culture in the newsroom [14,15]. Consequently, social media editors (SMEs) serve as critical intermediaries between digital audiences, newsrooms, and marketing departments [12]. Given their strategic positioning within the organizational hierarchy, SMEs are tasked with the continuous refinement of engagement strategies [15]. This process is increasingly facilitated by BI tools, which enable the systematic monitoring of decision outcomes and the iterative optimization of strategies to foster deeper audience engagement. Concurrently, the proliferation of COVID-19-related misinformation [16] has necessitated the integration of automated detection models [17], providing SMEs with the computational support required to identify and mitigate the spread of fraudulent content.

Furthermore, the research literature performed by Rejeb et al. [18] showed a lack of research on studying SM strategies [19], particularly in media [20,21]. Having recognized this gap and motivated by the increase in interest in SM strategy [12,22], especially in the media industry [23], this research focuses on the effectiveness of a visual content data-driven strategy used by the Portuguese publisher *Público*.

The bibliometric analysis presented by Rejeb et al. [18], regarding Instagram research, concludes that studies can be carried out using quantitative approaches to better understand the role of Instagram in influencing users' behaviour. *Moreover, inspired by the successful business case of financial times* [24], *the multi-dimensional RFV (recency, frequency, and volume)* role of Instagram in influencing users' behaviour. Moreover, in the recent research on Instagram use, the COVID-19 pandemic is one of the dominant topics [18]. However, none related to how to implement a data-driven strategy to avoid misinformation and promote constructive journalism [18] with visual content.

Furthermore, authors argue [7] that there are few studies on the approaches that organizations use to communicate information on Instagram, as there are some limitations, such as how the platform's algorithm works. However, there is a strong interest in understanding how to meet audiences' needs [25] and which types of content have more responses or engagement from the readers [7]. Moreover, recent researchers have tried to explore how audiences are responding to content on Instagram [25,26]. Thus, in this exploratory research, we aim to address four research questions (RQs) that are linked to the four main reader funnel steps presented in Figure 4. In the first step, ''attract reader,'' we would like to know whether there are new approaches and formats of Instagram posts that are more engaging? Which type of message is the most effective? (RQ1). Then, from a journalistic point of view, we aim to understand whether there is a constant/unbiased tone in caption posts that ensures authenticity and trustworthiness? (RQ2). Then, regarding the conversion into a website reader, we explore how a data-driven strategy can leverage website readers from Instagram? (RQ3). Finally, we aim to explore whether those readers present a high propensity to convert as paid subscribers, that is, are Instagram visits contributing to new online newspaper subscribers? (RQ4)

By answering these RQs, this study aims to contribute to the literature on visual communication, visual content data-driven strategy, and also to guide SMEs on how to effectively promote content that users engage with and consequently convert into subscribers. Drawing on data from external and internal databases, collected between January 2020 and September 2021, we analyse the strategy followed.

2. Literature review

2.1. Reader engagement and SM

Authors agree that reader engagement is a multi-dimensional phenomenon [27] encompassing emotional, cognitive, and behavioural involvement with media content [28–31]. Extensive research has further established a significant correlation between SM strategies and heightened levels of engagement [32,33]. Concurrently, recent studies highlight a paradigm shift in news production, characterized by an increased reliance on data analytics and visual SM content to facilitate multi-platform dissemination [15]. However, in the case of Instagram, literature is rather scarce, especially in the media industry [6]. By consulting the digital database Scopus, which is the largest abstract and citation database of peer-reviewed literature [34], the following query to articles' titles, abstracts, and keywords only returns 56 relevant documents (as of September 2025).

```
(TITLE-ABS-KEY (instagram AND (newspaper OR engagement OR online OR
publisher) AND news AND strategy))
```

Throughout the COVID-19 pandemic, SM has become the mainstream communication tool for content generation, dissemination, and consumption [5,11,18,35]. This dominance was largely driven by government-mandated non-pharmaceutical interventions, which restricted physical mobility and necessitated prolonged periods of confinement [36]. Despite its utility, SM platforms also facilitated the rapid proliferation of misinformation and biased information [5,37,38]. In fact, empirical evidence has established a significant correlation between the consumption of false

information and adverse psychological outcomes, including heightened mental distress, self-harm, and increased suicide rates [19,39,40]. Conversely, the literature suggests that constructive journalism serves as a vital mitigating factor, exerting a positive influence on public well-being during periods of crisis [23].

A systematic review of literature on health-related misinformation [41] presents the main psychological aspects that make the snowballing of misinformation occur. The authors concluded that potential interventions should act on cultivating critical thinking (Wang et al.). Furthermore, research indicates that during the COVID-19 pandemic, the use of media for information seeking was associated with increased loneliness and, in turn, decreased satisfaction with life [42]. Thus, journalism plays a crucial role in society and democracy by providing verified knowledge about current events [43] and creating social bonds in times of crisis [18], by adding awareness and distributing accurate content [44] on the Internet.

In the Portuguese context, news organizations proactively mitigated the spread of misinformation by suspending paywalls for critical COVID-19-related coverage [5]. The subsequent growth in the subscriber base [45] proves that readers pay for trustworthy journalism [8]. Consequently, SM platforms provide significant opportunities for expanding audience reach [46] and fostering reader advocacy [47]. By implementing data-driven strategies across the reader conversion funnel, publishers can leverage these platforms to stimulate subscription growth and secure long-term institutional loyalty [3].

2.2. Instagram and visual journalism

Studies have examined user-generated content (UGC) to explore determinants of user engagement on social networks by focusing more on text analysis [33] and less on visual content [9,23]. However, recent experiments suggest that including pictures in a post leads to greater levels of engagement [48]. Furthermore, visual solutions journalism [13] provides comprehensive coverage of an issue, showing problems and responses [49]. Moreover, the availability and low cost of mobile devices popularized videos as an example of virtual reality, promoting the rise of immersive journalism [13]. According to [32], ''vividness'' is related to how a post simulates several senses by incorporating sound, movement, and visual elements into the post or content. Research has studied vividness and its effects on news coverage of violence and war [23].

Instagram is a visual social network [9,18,50,51] and a multimedia sharing platform to use on mobile devices [18,19,52]. The user can upload images and videos [50], and it can use a range of options to improve content presentation, such as filters, GIFs, or captions [52]. Instagram is the fifth most preferred SM platform for sharing news [5]. Its SM penetration in Portugal was 69% [53]. Furthermore, Instagram was the fifth SM platform most used by users aged 16 to 64 years and verified the highest advertising reach compared with the total population aged more than 13 years (42%) [53]. In January 2020, 54% of Portuguese Instagram users were people aged 18 to 34 years. In 2021, this proportion was 50% [54], indicating that half of users are older than 35 years, the target with a higher propensity to buy an online newspaper subscription [55].

Instagram revealed a strong potential for brand awareness by spreading trustworthy content in a visual format [56] that reduces readers' uncertainty, especially in a crisis context [23].

2.3. Big data and analytics-driven journalism

Nowadays, companies are facing challenges to catch users' attention [24]. Due to the speed with which users interact, each second of time results in millions of reader interactions and huge volumes of data [24]. Editorial analytics have thus emerged as a critical component in media organizations [15], enabling audience-oriented editors to decode complex consumption environments and optimize strategic decision-making [14,57]. In fact, six main uses of analytics are identified by Lamot and Paulussen [14]: story placement, story packaging, story planning, story imitation, performance evaluation, and audience conception.

Moreover, SMEs are intermediaries between audiences, newsrooms, and marketing departments [12]. Thus, to maximize content mobile-first distribution, SMEs are allowed to observe internal and external data sources from historical to real-time data [15] through personalized BI (see Table 1). Analytic tools provide web analytics metrics to measure website performance and the overall story's engagement [15]. SMEs can identify missed opportunities and improve their strategy in future decisions. Furthermore, SMEs have access to real-time metrics that help to decide the content to be promoted by the platform according to content type and level of importance [14,58].

To quantify readers' engagement, most authors use attention-based metrics, such as average time per session or viewport time (i.e. time spent at each part of the article) [59]. Furthermore, the literature seems to agree that reader retention is related to how often a user reads (regularity) [60]. Consequently, this study analyses the frequency of sessions per user and pageviews per session as primary indicators of engagement (see Figure 5). Moreover, inspired by the successful

Table 1. Analytic tools and data available to optimize the decision-making process at Público.

Data sources		Historical data	Real-time data
Internal	Website	Web analytics dashboards for general evaluations and comparisons on a longer-term basis Reader's profile, that is, sociodemographic data, content consumption by day of the week, hour, author, or tag	Editorial Intelligence tool for real-time observational data about readers and website intra-day management
	Registered and subscribers	Main registration or subscription articles, tags, and authors Register or subscriber's profile	
External	Social networks Competitors	Content discovery and social monitoring platform	

business case of financial times [61], the multi-dimensional RFV engagement metric [62] was operationalized at Público in September 2020 to provide a more granular assessment of audience loyalty.

The widely applied RFM (recency, frequency, and monetary) model [63] was adapted to the media business context by replacing the monetary dimension with the number of articles visited, that is, the volume of content consumed. Thus, R measures how recently the consumer has visited the website in days, F is the number of website visits within a time period, and V indicates the total number of articles read in a period. The RFV scoring process, for quantifying reader behaviour, uses the quintile method [63]. Thus, an RFV value that goes from 111 to 555 is assigned to each reader. A total of 125 combinations ($5 \times 5 \times 5$) that divide the sample into 125 equal clusters. The readers who have the highest RFV scores are the most engaged, corresponding to readers who visited the website most recently, made more visits (highest frequency), and read many articles (highest volume) for 90 days.

To measure users' SM engagement in the platform [31], researchers study two types of metrics [48]. One related to the direct response, or interactions, with the posts, such as the number of likes or comments [32]. The second is related to the sharing and propagation of the published content that allows the user to recommend content to their followers. In this study, we will focus on interactions, such as the number of likes or comments and the number of video visualizations. As Instagram is a photo-sharing SM platform [9,48], we aim to examine how posts' characteristics (e.g. content, interactivity, and vividness) [32] impacted user engagement and the relation between post engagement and website traffic from Instagram users.

To answer the previously raised RQs (also presented in Figure 1), our study consisted of five distinct steps: editorial strategy understanding, data gathering and analysis, results achieved, discussion, and, finally, limitations and conclusions.

3. Methodology

3.1. Editorial and content strategy of Público

In the daily routine, SMEs monitor audience engagement by quantifying levels of attentiveness and interactivity to continuously improve and create more engaging and relevant content posts. This dynamic workflow is sustained by a multidisciplinary team integrating expertise in graphic design, content distribution, and data analytics elements dedicated to the empirical testing of innovative engagement strategies. Guided by real-time dashboard analytics and predefined key performance indicators (KPIs), the team systematically adapts creative approaches to align with the evolving behavioural patterns of Instagram users. Hence, the main goal in the digitalization and platformization of Público journalism was to raise readers' engagement along the reader funnel [3] (see Figure 2) by distributing the right content at the right time in the right format.

Thus, the researcher made available a central dashboard page (see Figure 2) to allow SMEs to easily find the main information to drive the decision-making process (real-time and historical data, respectively, shown in blue and green colour). Thus, depending on SME information needs, they can take real-time decisions based on what is happening now by using real-time dashboards, or what the trend was in the past by analysing historical dashboards. As an informed data-driven process, the team analyses the content produced by journalists and decides which stories can be promoted

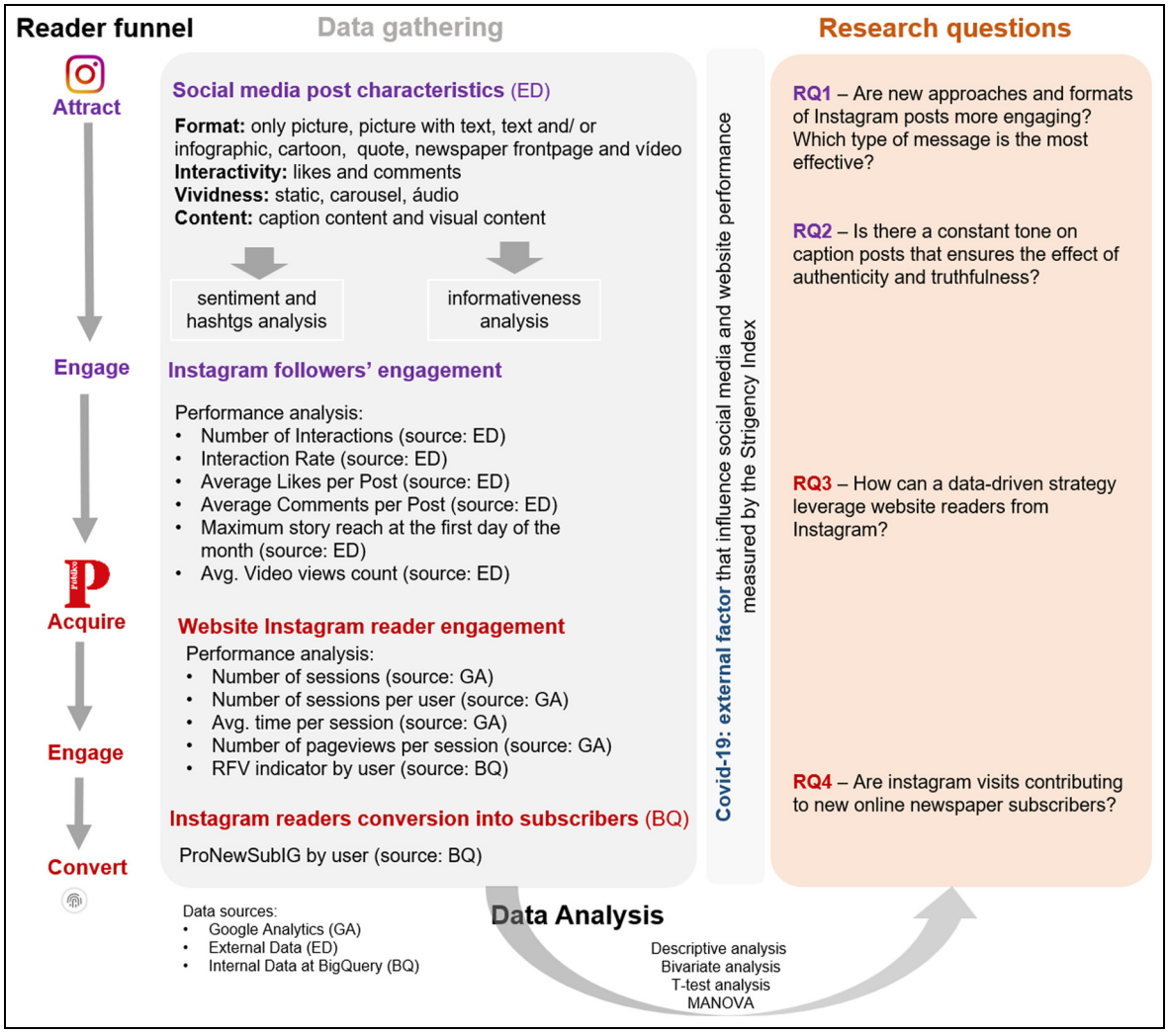


Figure 1. This figure delineates the operational variables and RQs mapped across the successive stages of the reader conversion funnel.
The framework's primary objective is the optimization of strategic touchpoints designed to facilitate the transition of audience traffic to the primary website. By utilizing Instagram as a high-visibility acquisition vector, the model illustrates the systematic transformation of social media impressions into sustained deep-site engagement, thereby nurturing the progression from casual readership to premium subscription.

on Instagram. Furthermore, SMEs do not influence content production induced by engagement metrics; they optimize content distribution.

The following sections delineate the iterative decision-making framework employed by SMEs, characterized by its responsiveness to real-time audience metrics and its commitment to delivering high-utility, empirical information. This strategic workflow is fundamentally anchored in the core institutional values of *Público*, ensuring that content distribution remains rigorous, authoritative, and intrinsically valuable. By synthesizing real-time behavioural analytics with traditional journalistic standards, the process bridges the gap between digital reactivity and editorial integrity.

- First, to take real-time decisions, SMEs combine information from the real-time analytics platform (see blue colour in Figure 2) and the social monitoring platform to monitor content performance across channels, devices, or user types, as well as to monitor and discover misinformation spread in SM that can be clarified with Público content. The goal is to optimize content distribution by channel, along with informing and clarifying the misinformation that is over-performing in other SM accounts. As an example, the content presented in Figure 3 shows a carousel post and a story that clarifies to readers about the hypothetical pandemic end.

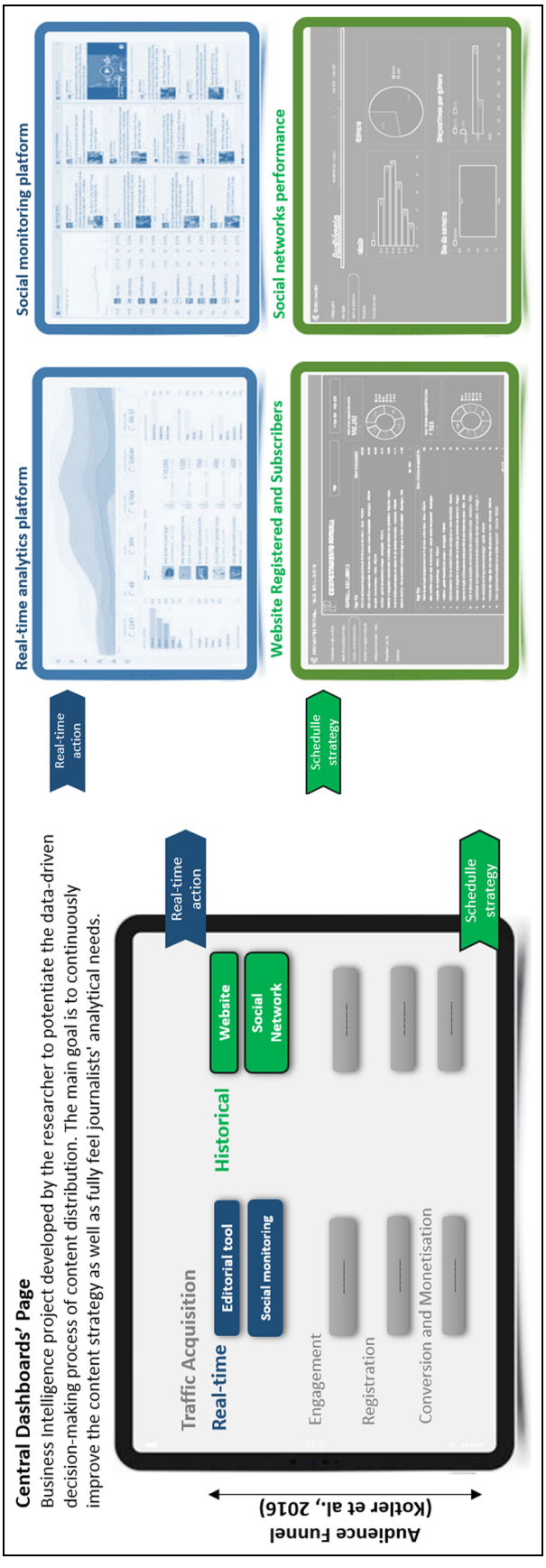


Figure 2. This schematic illustrates the synthesis of real-time and historical data streams from web and social media platforms, operationalized through a centralized Google Looker Studio (formerly Data Studio) ecosystem [64].
The interface utilizes a compartmentalized dashboard hierarchy, enabling SMEs to distinguish between immediate behavioural triggers and longitudinal trends across the reader conversion funnel. By centralizing these metrics, the system facilitates an empirical evaluation of editorial interventions and their channel-specific impact on audience progression.



Figure 3. Example of an informative quiz, posted in a story and in a carousel, about group immunity and the possible pandemic end, posted on 19 March 2021.

- Second, by identifying the content produced by the newsroom that needs to be posted on Instagram (i.e. content ideation), the team defines the design approach (content creation) and selects the right time for publishing. Furthermore, BI work from the researcher was performed to develop proper dashboards that provide historical data from different sources (see at Figure 2 in green colour we present the information flow related to historical data. Those pages provide top articles read by registered readers and subscribers, which allows the editor to see levels of engagement by tag or section). Furthermore, the dashboards allow SMEs to analyse previous decisions, to decide the content to schedule, and to promote content on Instagram according to followers' profiles as well as previous content performance. Dashboards developed by the researcher allow the comparison of behaviours between anonymous, registered, and newspaper subscribers. Thus, by customer type, SME can compare engagement KPI's across social media sources and devices in the green section at Figure 2.

As an example, evergreen content, that is, content that is consistently of readers' interest [64], is presented at the historical dashboard. By cross-checking the matters that had good performance by hour or day of the week from Instagram, and the evergreen content for this target, the SME decides the content to schedule on Instagram. As an example, in Figure 4, two types of content are presented, one related to Alzheimer's disease and the second about the bees' world. Both are multimedia formats produced by the newsroom that SMEs posted because usually Instagram followers engage with these formats and content.

- Finally, entertainment content, such as pictures or videos, is scheduled according to the best days and hours for this format.

The team also focuses on writing informative text in the posts, with an unbiased tone to ensure neutrality and steadiness [11], and a careful hashtag selection. Moreover, the team meets frequently with the design team to improve content ideation, creation, and planning, always focused on readers' profiles (i.e. readers' age and preferred tags). For SMEs, it becomes clear which audiences they want to focus on and the goals to achieve.

3.2. Data gathering

To evaluate the efficacy of the implemented strategy, this section delineates the multi-source data set synthesized for this study. To explore the dynamics of user engagement and conversion during the COVID-19 pandemic, Público provided



Figure 4. Example of content about Alzheimer's disease posted on 21 September 2020, and the "bee's world" posted on 6 June 2020.

access to primary data derived from four distinct sources: website data via Google Analytics (GA) [65], granular internal records retrieved from BigQuery (BQ), a Google-managed data warehouse [66], Instagram external data (ED) extracted through the InstLoadGram platform [67], and epidemiological data sourced from the *Our World in Data* repository [68] (as presented at Figure 4). Quantitative analysis and data processing were executed utilizing Microsoft Excel in conjunction with the R statistical computing environment [69].

As prior research suggested that there are several factors affecting the popularity of posts and subsequent engagement, such as the ease use of the technology, user motives and values, psychological factors, and post characteristics [32], the following post characteristics were collected (see Figure 1): Post date, content (caption, hashtags), interactivity (total number of likes, total number of comments, and total number of views by post), vividness (is carousel, is video, has audio), and video duration in seconds. Then, the number of hashtags per post and the number of words per caption were calculated. From Instagram, we also saved user profile characteristics such as age, gender, and city, as well as competitors' public data such as the number of followers.

Among the KPIs to analyse website users' engagement (see Figure 1), we highlight the time per session, typically expressed in seconds, an important KPI for publishers [57] that measures Instagram users' attention to the content on the website. Furthermore, at the individual level, we monitor the multi-dimensional RFV, an indicator of the reader's behaviour, implemented in September 2020, as well as the number of new subscribers who made at least one visit from

Instagram during the subscription month. That allows us to calculate the ProNewSubIG, which is the proportion of new subscribers with website visits from Instagram. These indicate how Instagram readers are converting into subscribers.

Besides, we added an external variable, the Stringency Index (see Figure 1) that "*is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest)*" [68]. After the initial data collection, we perform a descriptive analysis presented as follows.

4. Results

4.1. Message type and format analysis

In this section, we will focus on posts' characteristics by studying posts' interactivity, format, vividness, and content by post type, along with website users' engagement and conversion metrics. Message strategy and engagement effectiveness were examined. Microsoft Excel and the R statistical tool [69] were chosen to conduct the analysis.

Then, to analyse Instagram posts' effectiveness, all 2202 posts published at the Público Instagram account from January 2020 to September 2021 were scraped in November 2021. Content shared may be in a variety of formats, such as text, static or dynamic visuals, or a combination of all these [32]. Furthermore, each post can be published in a carousel-horizontal format, that is, by using multiple photographs or videos that can be viewed by swiping or clicking left. Thus, inspired by the study [48], posts were manually categorized into seven types: only picture, picture with text, text and/or infographic, cartoon, quote, newspaper frontpage, and video (see Table 2).

On the sample (see Table 2), less than half of the posts were static (948), most of which were in the form of a simple picture (47%) or a picture with text (42%). From all the content shared, 44% were carousels, 12% were videos with audio, and only 1% were videos without sound. In terms of post format, the most frequent (54%) was only a picture with 740 carousel posts and 449 static.

Despite the pandemic reality, in the first semester of 2020, Instagram engagement KPIs flattened (see Figure 5(b)). Thus, in the last trimester of 2020 (yellow box at Figure 5), a data-driven approach combined with an iterative editorial decision-making process was tested to address content readers' needs and preferences according to dashboards' information, coupled with design improvements. Depending on the content presentation, the level of interactivity differed. Thus, by experiments, the post's performance was continuously analysed to improve the next decision-making. As presented in Figure 5(a), more pictures with text and text infographics were posted. This resulted in a decrease in only picture posts (from 347 to 102, between 2020 and 2021, see Table 2).

In terms of interactivity, among the posts analysed, the mean number of likes was 4088 with a standard deviation of 4910 (see Table 2) as a result of different types of post content and vividness. Pictures with text posts verified the highest average number of likes (6625.3), followed by posts with author quotes (6441). Interestingly, the increase in average likes between 2020 and 2021 on pictures with text in carousel format (from 2292.1 to 8625.2), as well as text and/or infographic (from 2354 to 6809), reflects the importance of explaining the content visually and dynamically. This finding is consistent with the work of the author [49], which suggests that solutions-oriented photojournalism leads to higher user engagement [70].

Unsurprisingly, carousel posts revealed a high average number of likes by formats (see Table 2), except for the infographic. Researchers argue that vividness features in posts lead to higher total interaction [32]. In our sample, static text infographic posts verified the higher average likes in 2021 (11,789.6), because those posts were related to pandemic news and closures clarifications.

In what concerns comments, the average number of comments is higher on quotes (226.3), followed by pictures with text (180.2), proving that visual content with text promotes user reaction [70]. Furthermore, posts with author quotes on carousel present an average of 331 comments, revealing higher reader involvement in narrative content. These results are consistent with the work of the author [70].

To empirically examine the influence of the type of image content on reader engagement, a *t*-test was applied to compare the average likes by year (see Table 3). There is a statistically significant difference in the following formats: only picture, picture with text, cartoon, text infographic, and quote, proving the positive impact on interactivity as a result of the editorial process followed.

To analyse the effects of message type on the number of likes and comments. An MANOVA was conducted by using the post format as the independent variable. The overall MANOVA (Multivariate Analysis of Variance) tests of Pillai's are significant ($p < 0.0001$), suggesting that the number of likes and comments varied across post formats, supporting the hypothesis that different post formats generate different post effectiveness.

Table 2. Instagram post statistics by format considering vividness levels and interactivity metrics.

				Format								
				Only picture	Picture with text	Text infographic	Cartoon	Quote	Newspaper frontpage	Video		Total
										No audio	Has audio	
Vividness	Period	Interactivity	*N*	1198	556	93	61	11	15	19	259	**2202**
			%	54%	25%	4%	3%	0%	1%	1%	12%	**100%**
Static	2020		*N*	347	108	9	36	3	8			**511.0**
		Likes	Avg.	2539.7	3627.7	3715.2	1874.0	872.0	1206.8			**2712.8**
			SD	2075.8	3469.1	3592.1	1443.3	105.4	735.1			**2474.2**
		Comments	Avg.	49.6	102.9	122.7	56.1	45.0	13.9			**62.0**
			SD	55.9	109.1	119.1	155.0	30.5	17.5			**84.0**
	January–September 2021		*N*	102	291	19	21	1	3			**437**
		Likes	Avg.	5678.8	7018.2	**11,789.6**	3340.9	1143.0	2528.0			**6692.0**
			SD	6410.1	5326.6	11,228.9	2821.6		1737.2			**6012.4**
		Comments	Avg.	117.4	215.4	256.8	97.0	37.0	25.7			**186.9**
			SD	223.9	235.7	183.3	245.6		4.0			**235.0**
	Total		*N*	449	399	28	57	4	11			**948**
			%	47%	42%	3%	6%	0%	1%			**43%**
		Likes	Avg.	3252.8	6100.5	9194.3	2414.4	939.8	1567.1			**4547.1**
			SD	3784.8	5117.0	10,130.6	2157.5	160.5	1167.4			**4886.5**
		Comments	Avg.	65.0	185.0	213.7	71.2	43.0	17.1			**119.6**
			SD	120.5	214.9	175.1	192.2	25.2	15.8			**182.0**
Carousel	2020		*N*	512	15	32	4	1	4			**568**
		Likes	Avg.	2457.5	**2292.1**	**2354.4**	4670.3	2204.0	2326.3			**2461.6**
			SD	1831.8	2111.3	2470.8	6156.9		1095.7			**1926.4**
		Comments	Avg.	33.2	46.8	34.3	28.3	50.0	28.8			**33.6**
			SD	49.4	63.2	38.4	39.3		28.2			**49.0**
	January–September 2021		*N*	228	141	33		6				**408**
		Likes	Avg.	5822.8	**8625.2**	**6809.2**		10,815.5				**6944.5**
			SD	6608.3	7315.6	7787.6		4772.0				**7051.2**
		Comments	Avg.	72.1	183.1	138.9		377.8				**120.4**
			SD	124.7	255.3	224.0		177.6				**197.7**
	Total		*N*	740	156	65	4	7	4			**976**
			%	76%	16%	7%	0%	1%	0%			**44%**
		Likes	Avg.	**3495.4**	**8016.2**	4616.0	**4670.3**	**9585.3**	**2326.3**			**4335.6**
			SD	4260.5	7228.4	6190.1	6156.9	5437.9	1095.7			**5273.2**
		Comments	Avg.	45.2	170.0	87.4	28.3	331.0	28.8			**69.8**
			SD	82.4	246.7	169.0	39.3	204.1	28.2			**139.8**

(continued)

Table 2. (continued)

				Format								
				Only picture	Picture with text	Text infographic	Cartoon	Quote	Newspaper frontpage	Video		Total
										No audio	Has audio	
Vividness	Period	Interactivity	*N*	1198	556	93	61	11	15	19	259	**2202**
			%	54%	25%	4%	3%	0%	1%	1%	12%	**100%**
Audio	2020		*N*							14	158	**158**
		Likes	Avg.							869.9	1351.8	**1351.8**
			SD							852.8	1439.1	**1439.1**
		Comments	Avg.							8.9	39.2	**39.2**
			SD							9.8	55.2	**55.2**
		Count views	Avg.							16,842	23,969	**23,969**
			SD							12,171	18,164	**18,164**
	January–September 2021		N							5	101	**101**
		Likes	Avg.							648.6	2281.7	**2281.7**
			SD							318.7	3224.2	**3224.2**
		Comments	Avg.							10.0	76.8	**76.8**
			SD							9.3	155.4	**155.4**
		Count views	Avg.							14,351	39,095	**39,095**
			SD							6998	39,584	**39,584**
	Total		*N*							19	259	**278**
		Likes	Avg.							811.6	1714.4	**1652.7**
			SD							746.9	2344.4	**2282.0**
		Comments	Avg.							9.2	53.9	**50.8**
			SD							9.4	107.5	**104.4**
		Count views	Avg.							16,187	29,867	**28,932.5**
			SD							10,915	29,372	**28,692.4**
Total		*N*		1189	555	93	61	11	15	**278**		**2202**
		%		54%	25%	4%	3%	0%	1%	**13%**		**100%**
		Likes	Avg.	3383.6	6625.3	5994.4	2562.4	6441.5	1769.5	**1652.7**		**4087.9**
			SD	4078.8	5848.4	7825.2	2560.7	6064.3	1162.5	**2282.0**		**4910.6**
		Comments	Avg.	52.4	180.2	125.4	68.4	226.3	20.2	**50.8**		**88.9**
			SD	98.6	224.2	179.6	186.2	215.2	19.4	**104.4**		**158.2**

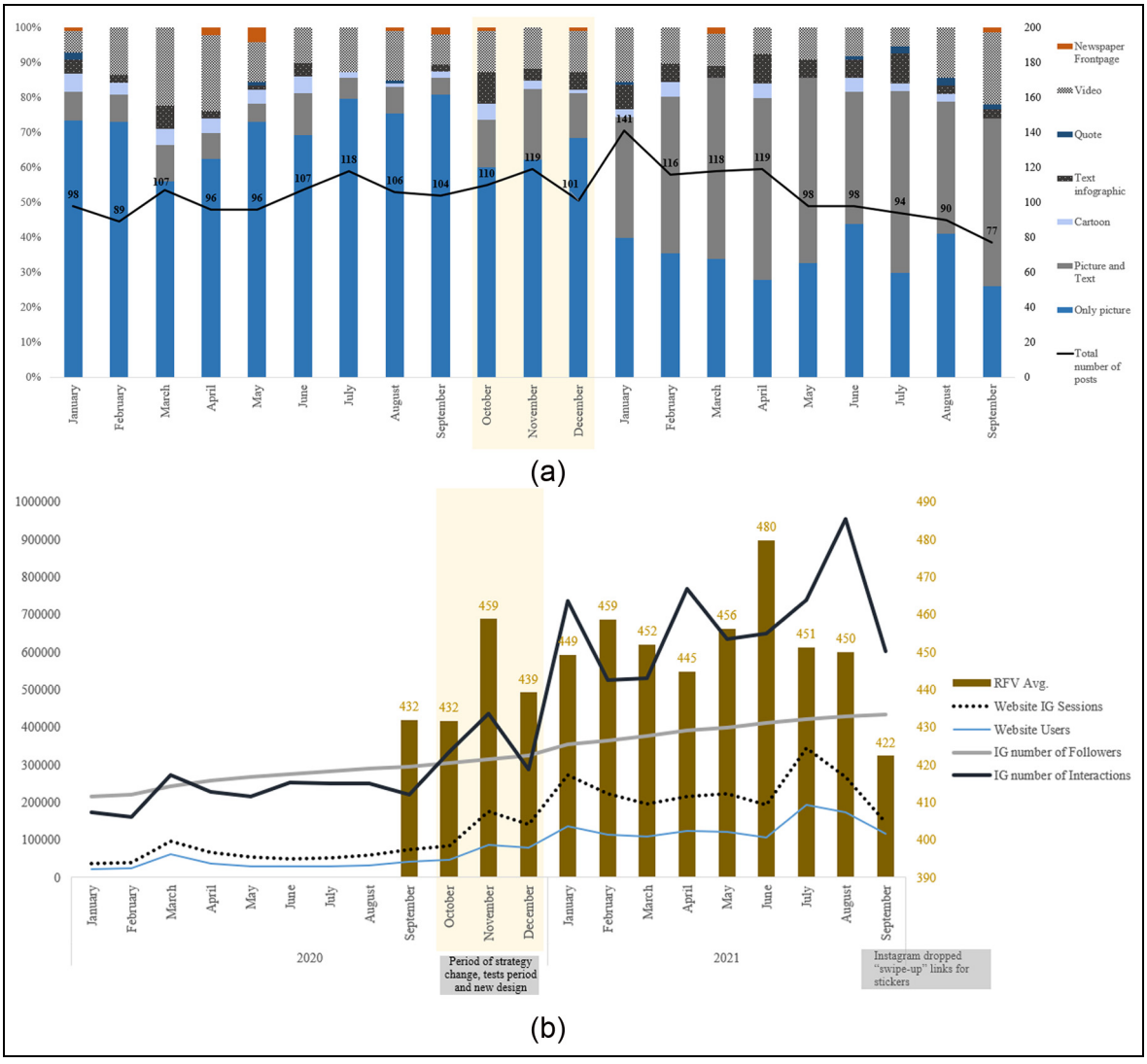


Figure 5. (a) Monthly posts by post category and (b) website number of users from IG, sessions, and, respectively, average RFV, plus Instagram followers and interactions during the timespan under study.

4.2. Text analysis

In terms of text content features, we consider the influence of caption content and visual content. Therefore, on the caption, we study text sentiment and text dimension (number of words and number of hashtags); on the image, we analyse informativeness, that is, if the image was edited by adding text or not.

The sentiment score was calculated through the sentimentR package using the open-source R statistical tool [69]. SentimentR was designed to calculate text polarity sentiment (TPS) at the sentence level and optionally aggregate by rows or grouping variable(s) [71]. For corpus handling and text preprocessing, we used the tm [72], the NLP [73], and the qdap packages [74]. Preprocessing involved punctuation, numbers, and white spaces removal, as well as Portuguese stopwords such as ''de'' or ''a.'' A score of 0 means that after the algorithm processed the sentence, the net emotional polarity was zero. This happens in two main scenarios:

Purely factual: the sentence contains no words found in the sentiment dictionaries (e.g. ''The table is made of wood'').

Perfect balance: the sentence contains positive and negative elements that cancel each other out perfectly (e.g. ''The movie was good, but the performance of the actor was bad'').

Table 3. *t*-Test: Two-sample assuming unequal variances.

t-Test: Two-sample assuming unequal variances	Only picture	Picture and text	Cartoon	Text infographic	Video	Quote	Newspaper frontpage
t Stat	−3636	−6356	−2749	−4924	−1875	−3128	−1105
P(T ⩽ t) two-tail	0.007	0.000	0.023	0.001	0.094	0.035	0.468
t Critical two-tailed	2306	2145	2262	2262	2262	2776	12,706

Decision rule: **if t Stat** < − **t Critical two** − **tail** or **if t Stat** > **t Critical two** − **tail** reject the null hypothesis.

In this study, the descriptive statistics indicate that captions tend to be neutral (TPSmedian = 0.000 and TPSmean = − 0.0036), ensuring an unbiased tone in text posts (RQ2) communication, which means that text posts are either devoid of emotion or perfectly balanced. The maximum TPS is 0.5897 related to a post about Martin Luther King published on 16 January 2020, and the minimum TPS (−0.8215) is related to a cartoon about Donald Trump and democracy published on 11 January 2021.

Furthermore, by calculating a term frequency-inverse matrix, the words with more relative importance among captions emerge. As expected, in the top 10, we found *Portugal, Lisboa, covid, pandemia, mundo (world)*, and *coronavirus*. Also, we note that words like *animais (animals), pessoas (people), ambiente (environment), and natureza (nature)* are also frequent, revealing the diversity of content published.

Moreover, the text dimension was studied by calculating the number of words and hashtags. The number of post words ranges between 6 and 284 words, with an average of 80. The number of hashtags varies between 0 and 20, with an average of 7 per post. Thus, the variable number of words was categorized into six classes and the variable number of hashtags into four classes, as presented in Table 4.

Among the 4149 hashtags, the authors also examine hashtag frequency and the engagement KPIs registered on the respective hashtags' posts. On the most used hashtags, we found #covid (644), #portugal (346), #pandemia (300), #coronavirus (177), #pandemic (160), and #animais (153). Between the hashtags with frequency higher than 50, we found #ambiente, #natureza, #ciencia, or #sociedade. However, the most used hashtags are not necessarily related to the hashtags from posts with high interactivity. Therefore, we calculate an engagement score that sums the normalized average number of likes per hashtag and average number of comments per hashtag, giving each one a weight of 50%. A hashtag with a high engagement score value indicates that, among the sample, this hashtag was used on posts that registered the highest average number of likes and average number of comments.

$$\text{engagement Instagram score} = \frac{X_{\text{likes}} - \overline{x}_{\text{likes}}}{s_{\text{likes}}} 0.5 + \frac{X_{\text{comments}} - \overline{x}_{\text{comments}}}{s_{\text{comments}}} 0.5 \quad (1)$$

For instance, #Tokyo presents the highest engagement score (3.21) in 2020, despite being the 48th hashtag most used. While, in 2021, #sociedade presented the highest score (4.55), and it was in the 45th place of the rank (see Figure 6). Furthermore, the year 2021 presents more hashtags with positive scores (24 versus 19 in 2020), indicating more engagement in a wide variety of matters, such as architecture, Joe Biden, tourism, history, or biodiversity.

In both years, sports posts conquer higher levels of engagement (#tokyo, #desporto, #futebol), as well as animal posts (#animais). As expected, hashtags related to the pandemic or related to the Portuguese actuality (#sociedade) are at the top of the most engaging posts and are also the most used. Interestingly, the hashtag #covid is the first one more written on both years; however, in 2020, the average number of likes and comments by post was significantly lower than the other hashtags; thus, the engagement score was less than zero. That does not happen in 2021, indicating that the decisions made to improve content presentation on Instagram, as well as to clarify pandemic information, resulted in more user involvement.

As presented in Table 4, the average number of comments on posts with edited text on the image (PwTextI) is three times higher than without text, and the average number of likes is more than double, indicating that the presence of more information induces more users' reactions. Interestingly, PwTextI with edited text on the image increased (194%), especially on posts with more than 101 words and less than 150 words (930% YoY).

Furthermore, by comparing the average likes, by groups of caption number of words, posts without text increase the number of likes as the number of words increases in the caption (from 2669 to 4306). While PwTextI presents values between 6172 and 7643 likes. In terms of hashtags, 467 of the 650 PwTextI present between five and nine hashtags. However, posts with more likes and comments present fewer than five hashtags, indicating that the user reaction is more related to the information on the image than the number of hashtags.

Table 4. Post's Instagram engagement metrics by caption number of words and number of hashtags.

		Post with text on the image (PwTextI) (more informativeness)				Post without text on the image				Post with text versus without text
Caption No. of words		2020	2021	Total	YoY	2020	2021	Total	YoY	Total var.
[5–50]	#	16	15	31	−6%	174	68	242	−61%	−87%
	Avg. likes per post	2581	13,042	7643	405%	2079	4177	2669	101%	186%
	Avg. comments per post	76	264	167	246%	52	81	60	56%	179%
[51–100]	#	125	218	343	74%	674	190	864	−72%	−60%
	Avg. likes per post	3294	7821	6172	137%	2238	5278	2906	136%	112%
	Avg. comments per post	86	191	153	123%	39	104	53	166%	188%
[101–150]	#	23	237	260	930%	228	186	414	−18%	−37%
	Avg. likes per post	3470	7023	6708	102%	2584	4840	3598	87%	86%
	Avg. comments per post	87	202	192	131%	33	77	53	133%	261%
[151–200]	#	1	13	14	1200%	5	18	23	260%	−39%
	Avg. likes per post	2887	7838	7485	172%	1836	4992	4306	172%	74%
	Avg. comments per post	68	247	234	263%	25	52	46	107%	407%
[201–250]	#		1	1		3	2	5	−33%	−80%
	Avg. likes per post		39,078	39,078		683	1329	941	95%	4051%
	Avg. comments per post		1282	1282		6	14	9	133%	13,835%
[251–350]	#		1	1		2	2	4	0%	−75%
	Avg. likes per post		1808	1808		1924	2049	1987	6%	−9%
	Avg. comments per post		28	28		57	43	50	−25%	−43%
Caption No. of hashtags										
[0–4]	#	33	91	124	176%	199	134	333	−33%	−63%
	Avg. likes per post	3461	8362	7058	142%	1997	3581	2634	79%	168%
	Avg. comments per post	100	260	217	160%	36	80	54	123%	305%
[5–9]	#	107	360	467	236%	673	296	969	−56%	−52%
	Avg. likes per post	3243	7710	6687	138%	2366	5350	3277	126%	104%
	Avg. comments per post	84	194	169	130%	42	88	56	109%	201%
[10–14]	#	25	34	59	36%	196	33	229	−83%	−74%
	Avg. likes per post	2982	5037	4166	69%	2288	5899	2808	158%	48%
	Avg. comments per post	69	135	107	94%	36	111	47	211%	129%
[15–21]	#					18	3	21	−83%	−100%
	Avg. likes per post					2028	8510	2954	320%	−100%
	Avg. comments per post					35	69	40	97%	**−100%**
Total	#	**165**	**485**	**650**	**194%**	**1086**	**466**	**1552**	−57%	**−58%**
	Avg. likes per post	**3247**	**7645**	**6529**	**135%**	**2278**	**4900**	**3066**	115%	**113%**
	Avg. comments per post	**85**	**202**	**172**	**138%**	**40**	**87**	**54**	120%	**220%**

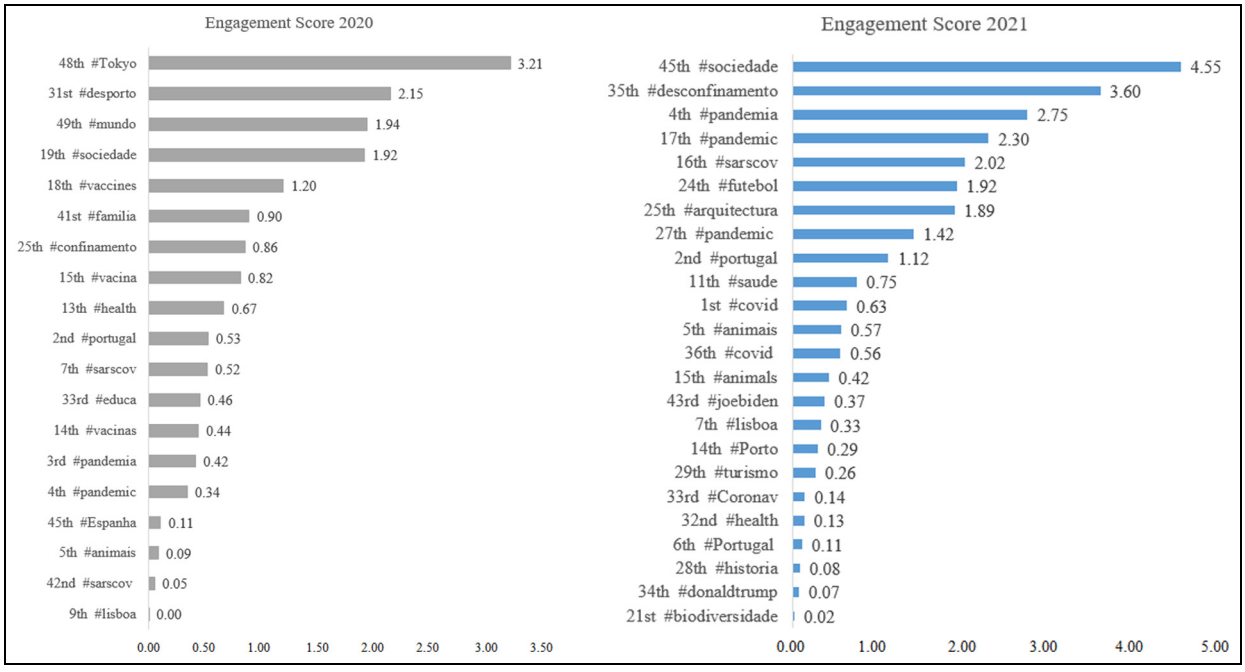


Figure 6. Hashtags with positive engagement score by year and respective rank position.
For example, in 2020, the hashtag #portugal was the second most used with a score of 0.53.

4.3. Engagement and conversion between periods

To examine the impact of the new strategy followed in 2021, Table 5 presents the *t*-test on the metrics between 2020 and 2021. There is a statistically significant difference between the average number of sessions by month between periods, because the sessions increased by over 17% monthly in 2021. The monthly average of the number of interactions, interaction rate, likes per post, comments per post, and stories reach also presented a statistically significant difference between the two years.

Furthermore, as presented in Figure 5(b), since September 2020, there was an increase in followers (3% monthly on average), and the average number of monthly website visits more than doubled. The monthly number of interactions registered an increase of more than 265% between 2020 and 2021. Moreover, we also found a significant increase in the proportion of new subscribers that came to the website from Instagram (3% in 2021). Besides the monthly number of posts (see Figure 5), the number of words and hashtags per post remained relatively stable across both years, at approximately 88 words and 7 hashtags per post.

Meanwhile, the variables visits per user and average time session do not present a significant difference between years, indicating that, despite the peak in sessions, these users tend to keep the levels of engagement during website visits. In fact, the RFV indicator, since September 2020, presented an average value of 448 (see Figure 5(b)) with a standard deviation of 14.7. Thus, the results suggested that the strategy followed attracted an engaged audience.

4.4. Correlation analysis

According to the author [3], across the five stages of the customer funnel, the goal is to convert the maximum number of anonymous users into highly loyal customers [2]. Low levels of conversion rates result in low customer attraction, low curiosity, low commitment, and low affinity [3]. Thus, this section aims to analyse the relationship between the variables under study.

The correlation coefficient is a statistic that measures the strength of the relationship between two variables [75]. The Pearson correlation coefficient describes the proportion of the total variance in the observed data that can be explained by a linear model of the variables under study [76]. It takes values between -1 and 1. To interpret those values, labelling systems coupled with the statistical significance of the correlation coefficient allow us to define the kind of relationship that exists between variables [75]. The p value indicates the probability that the data would be inconsistent with the hypothesis. Thus, for a confidence level of 95%, when p value $\leqslant 0.05$, the correlation is significant [76].

Table 5. *t*-Test: Two-sample assuming unequal variances.

Variables under study	Statistics	2020	2021	t Stat	P(T⩽t) two-tail	t Critical two-tailed
Sessions	Avg.	34,922.33	167,715.44	−6.815	0.000	2.201
	SD	26,668.02	53,700.67			
Visits/user	Avg.	1.76	1.75	0.048	0.963	2.201
	SD	0.11	0.23			
Avg. time session (s)	Avg.	122.50	108.89	1.691	0.115	2.160
	SD	13.84	20.96			
PV/session	Avg.	3.53	2.30	4.586	0.000	2.093
	SD	0.74	0.48			
Interactions	Avg.	257,175.42	682,528.33	−8.569	0.000	2.179
	SD	73,541.59	134,600.04			
Interaction rate	Avg.	0.89	1.66	−5.871	0.000	2.262
	SD	0.12	0.38			
Avg. likes per post	Avg.	2099.87	6001.57	−5.681	0.000	2.262
	SD	622.97	1 988.49			
Avg. comments per post	Avg.	47.74	150.72	−4.353	0.002	2.306
	SD	14.22	69.90			
Maximum stories Reach on the first day of the month	Avg.	11,687.58	44,534.89	−3.424	0.009	2.306
	SD	4916.43	28,463.62			
ProNewSubIG	Avg.	0.01	0.03	−4.147	0.003	2.306
	SD	0.00	0.02			

Decision rule: **if t Stat**< − **t Critical two** − **tail** or **if t Stat**>**t Critical two** − **tail** reject the null hypothesis.

As presented in Table 6, the correlation matrix suggested that the variables – sessions, interactions, interaction rate, average likes per post, average comments per post, and ProNewSubIG – were significantly positively correlated. Furthermore, the number of followers increased along with the number of interactions; the interaction rate is significantly correlated with all the variables except the Stringency Index. Therefore, across the conversion funnel, as the number of Instagram followers increases, the number of readers on the website increases, and consequently, the subscription conversions increase.

4.5. Competitors analysis

To analyse Público's market position, Table 7 presents a comparison between market players.

In January 2020, Público already had more followers than its competitors. Through the strategy followed, Público increased 102% the number of followers and 251% the number of monthly interactions, leading the Portuguese ranking on this platform.

5. Discussion and RQ analysis

This addresses four primary RQs, as illustrated in Figure 1. Seeking to answer RQ1, previous analysis demonstrated that as the number of informative posts increases, and the subsequent growth of followers and interactions. This aligns with previous scholarship [49,70] suggesting that narrative engagement plays an important role involving audiences [24]. Notably, the follower base expanded significantly even while the monthly post volume remained constant, signifying a high degree of editorial efficacy in content selection. Moreover, the interaction rate and the maximum story reach on the first day of the month more than tripled during the observation period. Statistical validation via *t*-tests confirmed a significant positive difference in average likes and comments between the 2 years analysed. Consequently, the integration of textual information within visual formats serves as a critical mechanism for stimulating reader interest – a finding consistent with the empirical work of Aydin [32].

To answer RQ2, a text analysis was performed. The wide range of hashtags with a high engagement score indicates that there was a daily understanding of the reader's information needs. Furthermore, sentiment analysis demonstrated that the TPS tend to be neutral across the timespan, that is, the text is either devoid of emotion or perfectly balanced. Thus, engagement increased across the funnel while maintaining a journalistic ethical perspective. The content remained objective and fact-based, and was, in most cases, purely descriptive.

Table 6. Pearson correlation between variables.

KPIs		SpU	TpS	PpS	I	IR	LpP	CpP	Max. story	PropNewSubIG	SI
Website KPIs	Sessions	0.099	−0.356	−0.730 (**)	**0.869 (**)**	0.798 (**)	**0.822 (**)**	**0.807 (**)**	0.464 (*)	0.727 (**)	0.269
	Session per User (SpU)	1	0.417	−0.028	0.000	−0.227	−0.208	−0.168	−0.520 (*)	−0.281	0.470 (*)
	Avg. time per session (TpS)		1	0.269	−0.449 (*)	−0.612 (**)	−0.575 (**)	−0.542 (*)	−0.391	−0.720 (**)	0.338
	Pageview per session (PpS)			1	−0.747 (**)	−0.721 (**)	−0.752 (**)	−0.726 (**)	−0.527 (*)	−0.691 (**)	−0.416
Instagram KPIs	Interactions (I)				1	**0.949 (**)**	**0.945 (**)**	**0.863 (**)**	0.690 (*)	**0.879 (**)**	0.182
	Interaction rate (IR)					1	**0.945 (**)**	**0.863 (**)**	0.690 (*)	**0.879 (**)**	0.18
	Avg. likes per post (LpP)						1	**0.948 (**)**	0.690 (*)	**0.928 (**)**	0.193
	Avg. comments per post (Cp)							1	0.625 (*)	**0.904 (**)**	0.169
	Max. story								1	0.614 (**)	0.132
	PropNewSubIG									1	0.111
	Stringency Index (SI)										1

(*) correlation is significant at the level 0.05; (**) correlation is significant at the level 0.01 (two-tailed).

Table 7. Followers and monthly interactions comparison between Público and its competitors.

Portuguese publishers	January 2020		Comparison between September 2021 and January 2020	
	Público's number of followers versus competitors	Monthly interactions	Number of followers	Monthly interactions
Público	–	**171,686**	+ **102%**	+ **251%**
Expresso	49%	72,458	+ 83%	+ 105%
JN	41%	95,822	+ 73%	−39%
Observador	58%	57,054	+ 65%	+ 37%
Correio da Manhã	99%	21,287	+ 44%	−74%
Diário de Notícias	287%	6231	+ 51%	+ 949%

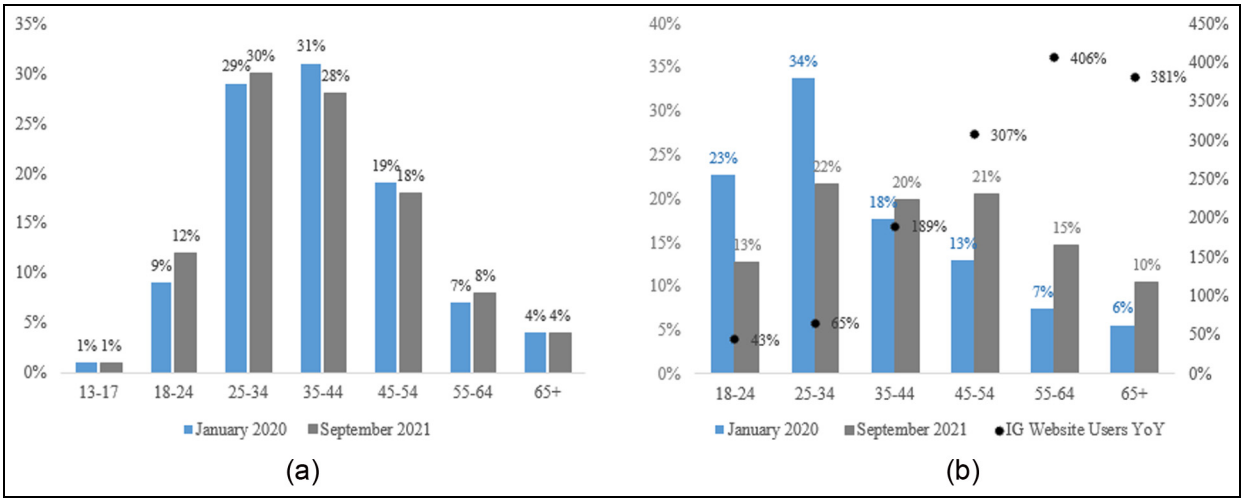


Figure 7. Instagram followers by age (a) and website users from Instagram by age (b).

Some authors defend that SM affects the way news is produced. However, Público SME's approach prioritizes the strategic curation of existing editorial content to maximize engagement among Instagram audiences [37]. This curation process is simultaneously informed by an institutional commitment to mitigating misinformation. By implementing a data-driven framework, the team was able to systematically identify and prioritize newsroom output that aligned with audience interests. The efficacy of this strategy is evidenced by the 2021 performance metrics: within the first 9 months, sessions originating from Instagram surpassed the total volume of the preceding year by 23%.

Seeking to investigate RQ3, a correlation analysis was performed. The low Pearson correlation between the Stringency Index and the number of sessions indicate that the increase in website sessions is not directly related to the strictness of lockdown. Thus, the increase in website sessions was a result of the efforts to optimize the SM strategy guided by data. Moreover, the strategic selection of content, the careful choice of content formats (detailed at the editorial strategy understanding section), drove followers who kept engaged with posts, as the interaction rate increased and consequently drove readers to the website. Despite concerns about glorifying analytic tools [15], this research aims to provide an example of how a data-driven SM strategy can keep journalism principles and inform readers who consequently become subscribers.

On the contrary, in terms of acquiring readers with more propensity to subscribe (RQ4), an increase of more than 300% on readers aged 45 years and more was verified, as a result of a 156% growth in website sessions from Instagram. In January 2020, the majority (60%) of Público Instagram followers fell in the age group 25–44 years and 30% had more than 44 years (see Figure 7(a)). When comparing January 2020 and September 2021, the distribution of the followers by age did not present significant differences. However, when we explore the users on the website that come from Instagram, the distribution is quite different (see Figure 7(b)), that is, 46% of the readers were more than 44 years. Interestingly, the daily decision-making process increased the right target of readers to subscribe as the historical dashboards are prepared to understand content consumption by hour, day, age, etc.

Besides, there is a strong positive correlation between PropNewSubIG, sessions, and interactions, indicating that the Público Instagram audience is reading more on the website, and it is reading content that induces the subscription purchase. Furthermore, Instagram readers on the website maintained the number of sessions per visit as the RFV engagement metric, revealing the assertiveness in content selection. Thus, answering positively to RQ4, the proportion of new monthly subscribers that visit the website from Instagram increased from 0.6% on January 2020 to 5.6% on August 2021. While some authors agree that SM contribution seems underwhelming [77], others agree that it can boost website engagement [31]. This research provides evidence that SMEs can apply a subscription data-driven strategy in SM that increases reader engagement without affecting journalistic output.

6. Conclusion and limitations

The diversification of journalistic distribution channels marks a structural transition from single-channel dissemination to integrated multichannel communication strategies [1,13,58]. This empirical study evaluates the impact of a SM content strategy that synthesizes reader engagement with data-driven methodologies. By leveraging high RFM (RFV) value content to drive web traffic, the strategy facilitates the conversion of casual visitors into subscribers. The findings demonstrate that users like and follow pages that provide them with useful and timely information in a convenient way [32]. Furthermore, text analysis revealed a predominantly neutral sentiment in social media output, suggesting that SMEs successfully eschewed ''clickbait'' tactics in favour of informational integrity [11]. Moreover, the implementation of an iterative testing and analysis methodology resulted in a measurable increase in engagement metrics across the reader conversion funnel. Thus, this paper shows how SMEs navigate the complexities of digital content delivery and adapt to evolving user behaviours and platform algorithmic updates. Furthermore, the BI framework developed herein provided the newsroom with actionable insights, ultimately enabling *Público*'s SMEs to realize their strategic objectives.

Indeed, the data-driven strategy leverages data analytics in four main fronts: at Instagram, to determine optimal content and timing to post, since follower habits continuously change during the pandemic; at the website, to determine the best content to post and attract users to the website with high propensity to subscribe; at real-time decision-making to inform breaking news and update latest top stories, and finally, on misinformation monitor to determine the content that need to be clarified.

Moreover, the proposed methodology and strategic framework demonstrate significant potential for scalability across diverse business contexts and communication industries. By integrating a culture of continuous experimentation with a robust BI strategy, content producers can systematically expand their audience reach and catalyse higher levels of engagement. This iterative approach facilitates the optimization of content performance through empirical validation, offering a replicable model for organizations navigating the increasingly volatile digital landscape.

Despite this paper identifying several working tactics and successful patterns, the study was not immune to limitations. This study only considered data from one Portuguese online newspaper between January 2020 and September 2021. Besides, other content information could be used to analyse post-performance, such as colourfulness or the presence of human faces. Furthermore, the Instagram platform is always evolving by changing features and algorithms. Thus, it could be interesting to analyse the evolution of the data-driven strategy to face Instagram changes and to keep growing engagement across the funnel.

Finally, SM strategy is dynamic and must be an ongoing task. Thus, future research could explore visual content SM strategy from other news media organizations in different countries.

ORCID iD

Elizabeth Fernandes https://orcid.org/0000-0002-2358-1229

Funding

The work by E. Fernandes and S. Moro was partially supported by national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT), ISTAR-Iscte Pluriannual Projects UID/04466/2025 (https://doi.org/10.54499/UID/04466/2025).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statements

The data sets analysed during this study are not publicly available due confidential company data by Público.

References

[1] Fernandes E, Moro S and Cortez P. Data science, machine learning and big data in digital journalism: a survey of state-of-the-art, challenges and opportunities. *Expert Syst Appl* 2023; 221: 119795.

[2] Karimi S, Shakery A and Verma R. Online news media website ranking using user-generated content. *J Inf Sci* 2021; 47: 340–358.

[3] Kotler P, Kartajaya H and Setiawan I. *Marketing 4.0: moving from traditional to digital*. John Wiley & Sons, 2016.

[4] Arrese Á. From Gratis to Paywalls: a brief history of a retro-innovation in the press's business. *J Stud* 2016; 17: 1051–1067.

[5] Newman N. *Digital news report: journalism, media, and technology trends and predictions 2022*. Reuters Institute, 2022, p.46.

[6] Cordeiro DF, Vállez M, Font-Julian CI, et al. Instagram engagement and content strategies of US and UK legacy media: a quantitative analysis of five leading news outlets. *J Media* 2025; 6: 89. https://doi.org/10.3390/journalmedia6020089

[7] Saks J and Hopkins A. Insta-news: how the New York Times and Washington post use Instagram. *Journal Pract* 2026; 20: 586–603.

[8] Rußell R, Berger B, Stich L, et al. Monetizing online content: digital paywall design and configuration. *Bus Inf Syst Eng* 2020; 62: 253–260.

[9] Araujo T, Lock I and Van de Velde B. Automated Visual Content Analysis (AVCA) in communication research: a protocol for large scale image classification with pre-trained computer vision models. *Commun Methods Meas* 2020; 14: 239–265.

[10] Ohme J, Araujo T, Boeschoten L, et al. Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Commun Methods Meas* 2024; 18(2): 124–141. https://doi.org/10.1080/19312458.2023.2181319

[11] Denisova A. Viral journalism. Strategy, tactics and limitations of the fast spread of content on social media: case study of the United Kingdom quality publications. *Journalism* 2023; 24(9): 1919–1937. https://doi.org/10.1177/14648849221077749

[12] Neilson T and Gibson TA. Social media editors and the audience funnel: tensions between commercial pressures and professional norms in the data-saturated newsroom. *Digit Journal* 2022; 10: 556–578.

[13] Bujić M, Salminen M, Macey J, et al. ''Empathy machine'': how virtual reality affects human rights attitudes. *Internet Res* 2020; 30: 1407–1425.

[14] Lamot K and Paulussen S. Six uses of analytics: digital editors' perceptions of audience analytics in the newsroom. *Journal Pract* 2020; 14: 358–373.

[15] Moyo D, Mare A and Matsilele T. Analytics-driven journalism? Editorial metrics and the reconfiguration of online news production practices in African newsrooms. *Digit Journal* 2019; 7: 490–506.

[16] Pampapura Madali N, Alsaid M and Hawamdeh S. The impact of social noise on social media and the original intended message: BLM as a case study. *J Inf Sci* 2024; 50: 89–103.

[17] Singh M, Ahmed J, Alam M, et al. A comprehensive review on automatic detection of fake news on social media. *Multimed Tools Appl* 2023; 83: 47319–47352.

[18] Rejeb A, Rejeb K, Abdollahi A, et al. The big picture on Instagram research: insights from a bibliometric analysis. *Telemat Inform* 2022; 73: 101876. https://doi.org/10.1016/j.tele.2022.101876

[19] Huang SL and Chang CY. Understanding how people select social networking services: media trait, social influences and situational factors. *Inf Manag* 57: 103323. https://doi.org/10.1016/j.im.2020.103323

[20] Hermida A and Mellado C. Dimensions of social media logics: mapping forms of journalistic norms and practices on Twitter and Instagram dimensions of social media logics: mapping forms of journalistic norms and practices on Twitter. *Digit Journal* 2020; 8: 864–884.

[21] Larsson AO. The news user on social media a comparative study of interacting with media organizations on Facebook and Instagram. *J Stud* 2018; 19: 2225–2242.

[22] Yu X, Yuan C, Kim J, et al. A new form of brand experience in online social networks: an empirical analysis. *J Bus Res* 2021; 130: 426–435.

[23] Overgaard CSB. Constructive journalism in the face of a crisis: the effects of social media news updates about COVID-19. *J Stud* 2021; 22: 1875–1893.

[24] Karampournioti E and Wiedmann KP. Storytelling in online shops: the impacts on explicit and implicit user experience, brand perceptions and behavioral intention. *Internet Res* 2021; 32: 228–259.

[25] Kallio S-M and Maenpaa J. Visuality as an affordance on Instagram news production. *Digit Journal*. Epub ahead of print 11 February 2025. https://doi.org/10.1080/21670811.2025.2462558

[26] Anter L and Kümpel AS. Young adults' information needs, use, and understanding in the context of Instagram: a multi-method study. *Digit Journal* 2025; 13: 1112–1130.

[27] Steensen S, Ferrer-Conill R and Peters C. (Against a) theory of audience engagement with news. *J Stud* 2020; 20: 1–19.

[28] Attfield S, Kazai G and Lalmas M. Towards a science of user engagement (position paper). In: *WSDM workshop on user modelling for web applications*, https://www.researchgate.net/profile/Benjamin-Piwowarski/publication/228542640_Towards_a_science_of_user_engagement_Position_Paper/links/0912f506997362eb8c000000/Towards-a-science-of-user-engagement-Position-Paper.pdf

[29] Ksiazek TB, Peer L and Lessard K. User engagement with online news: conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media Soc* 2016; 18: 502–520.
[30] Mersey RD, Malthouse EC and Calder BJ. Engagement with online media. *J Media Bus Stud* 2010; 7: 39–56.
[31] Manosevitch I and Tenenboim O. The multifaceted role of user-generated content in news websites: an analytical framework. *Digit Journal* 2017; 5: 731–752.
[32] Aydin G. Social media engagement and organic post effectiveness: a roadmap for increasing the effectiveness of social media use in hospitality industry. *J Hosp Mark Manag* 2020; 29: 1–21.
[33] Shahbaznezhad H, Dolan R and Rashidirad M. The role of social media content format and platform in users' engagement behavior. *J Interact Mark* 2021; 53: 47–65.
[34] Ballew B. Elsevier's Scopus® database. *J Electron Resour Med Libr* 2009; 6: 245–252.
[35] Volkmer I. Social media & COVID-19: a global study of digital crisis interaction among Gen Z and Millennials, 2021, https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/3958684/Volkmer-Social-Media-and-COVID.pdf
[36] Piccinelli S, Moro S and Rita P. Air-travelers' concerns emerging from online comments during the COVID-19 outbreak. *Tour Manag* 2021; 85: 104313.
[37] Cage J, Herve N and Mazoyer B. Social media and newsroom production decisions. *SSRN Electron J* 2020; 2020: 3663899. https://doi.org/10.2139/ssrn.3663899
[38] Apuke OD and Omar B. Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telemat Inform* 2021; 56: 101475. https://doi.org/10.1016/j.tele.2020.101475
[39] Rosenberg H, Syed S and Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Can J Emerg Med* 2020; 22: 418–421.
[40] Cheikh-Ammar M. The bittersweet escape to information technology: an investigation of the stress paradox of social network sites. *Inf Manag* 2020; 57: 103368. https://doi.org/10.1016/j.im.2020.103368
[41] Wang Y, McKee M, Torbica A, et al. Systematic literature review on the spread of health-related misinformation on social media. *Soc Sci Med* 2019; 240: 112552.
[42] Choi M and Choung H. Mediated communication matters during the COVID-19 pandemic: the use of interpersonal and masspersonal media and psychological well-being. *J Soc Pers Relat* 2021; 38: 2397–2418.
[43] Carlson M. *Journalistic authority: legitimating news in the digital era*. Columbia University Press, 2017.
[44] Perreault MF and Perreault GP. Journalists on COVID-19 journalism: communication ecology of pandemic reporting. *Am Behav Sci* 2021; 65: 976–991.
[45] Grzegorz P. *News subscriptions in the age of Coronavirus* (INMA Readers First Iniciative). International News Media Association, 2021.
[46] Nielsen RK and Ganter SA. Dealing with digital intermediaries: a case study of the relations between publishers and platforms. *New Media Soc* 2018; 20: 1600–1617.
[47] Choi TR and Sung Y. Instagram versus Snapchat: self-expression and privacy concern on social media. *Telemat Inform* 2018; 35: 2289–2298.
[48] Li Xie. Is a picture worth a thousand words? An empirical study of image content and social media engagement. *J Mark Res* 2020; 57: 1–19.
[49] Midberry J and Dahmen NS. Visual solutions journalism: a theoretical framework. *Journal Pract* 2020; 14: 1159–1178.
[50] Kang J, Yoon J, Park E, et al. ''Why tag me?'' Detecting motivations of comment tagging in Instagram. *Expert Syst Appl* 2022; 202: 117171. https://doi.org/10.1016/j.eswa.2022.117171
[51] Ghalavand H, Panahi S, Sedghi S, et al. Professional benefits and challenges of health information documentation on Instagram for Iranian physicians. *J Inf Sci* 2024; 50(3): 557–567. https://doi.org/10.1177/01655515221097860
[52] Jarreau P, Dahmen N and Jones E. Instagram and the science museum: a missed opportunity for public engagement. *J Sci Commun* 2019; 8: 55.
[53] Hootsuite. Digital 2020: Portugal, all the data, trends, and insights you need to help you understand how people use the internet, mobile, social media, and ecommerce, 2020, https://datareportal.com/digital-in-portugal.
[54] NapoleonCat Stats, W. Instagram users in Portugal, https://napoleoncat.com/stats/instagram-users-in-portugal (2021, accessed 7 November 2021).
[55] Goyanes M. Why do citizens pay for online political news and public affairs? Socio-psychological antecedents of local news paying behaviour why do citizens pay for online political news and public paying behaviour. *J Stud* 2020; 21: 547–563.
[56] Vraga EK, Kim SC, Cook J, et al. Testing the effectiveness of correction placement and type on Instagram. *Int J Press Polit* 2020; 25: 632–652.
[57] Hendrickx J, Montero E, Ranaivoson H, et al. Becoming the data-informed newsroom? The promotion of audience metrics in the newsroom and journalists' interactions with them. *Digit Journal* 2021; 9: 427–442.
[58] Neuberger C, Nuernbergk C and Langenohl S. Journalism as multichannel communication: a newsroom survey on the multiple uses of social media. *J Stud* 2019; 20: 1260–1280.
[59] Lagun D and Lalmas M. Understanding and measuring user engagement and attention in online news reading. In: *WSDM 2016—proceedings of the 9th ACM international conference on web search and data mining*, San Francisco, CA, 22–25 February 2016, pp.113–122. Association for Computing Machinery.

[60] Davoudi H. User acquisition and engagement in digital news media (issue December), 2018, https://yorkspace.library.yorku.ca/
[61] FTStrategies. How the financial times brought data into the newsroom. *FTStrategies*, 2022, https://www.ftstrategies.com/en-gb/insights/how-the-financial-times-brought-data-into-the-newsroom/
[62] Fernandes E, Moro S and Cortez P. A data-driven approach to improve online consumer subscriptions by combining data visualization and machine learning methods. *Int J Consum Stud* 2024; 48: e13030. https://doi.org/10.1111/ijcs.13030
[63] Ernawati E, Baharin SSK and Kasmin F. A review of data mining methods in RFM-based customer segmentation. *J Phys Conf Ser* 2021; 1869: 012085. https://doi.org/10.1088/1742-6596/1869/1/012085
[64] Liao Y, Wang S, Han EH, et al. Characterization and early detection of evergreen news articles. In: *ECML/PKDD*, Würzburg, Germany, 16–20 September 2019, pp.552–568. Springer.
[65] Semerádová T and Weinlich P. Using Google analytics to examine the website traffic. In: Semerádová T and Weinlich P (eds) *Website quality and shopping behaviour quantitative and qualitative evidence*. Springer, 2020, pp.91–112.
[66] Bisong E. Google bigquery. In: Bisong E (ed.) *Building machine learning and deep learning models on Google cloud platform*. Apress, 2019, pp.485–517.
[67] Zimnitskiy I. InstLoadGram, https://www.instaloadgram.com/ (2021, accessed 31 November 2021).
[68] Global Change Data Lab. Our world data, 2021, https://ourworldindata.org/
[69] Cortez P. *Modern optimization with R*. 2nd ed. Springer, https://pcortez.dsi.uminho.pt/mor-book#h.7ak9plkw1435 (2021, accessed 22 April 2023).
[70] Dahmen NS, Thier K and Walth B. Creating engagement with solutions visuals: testing the effects of problem-oriented versus solution-oriented photojournalism. *Vis Commun* 2021; 20: 271–288.
[71] True L. Package 'sentimentr', 2018, https://cran.r-project.org/web/packages/sentimentr/
[72] Feinerer I. Introduction to the tm package text mining in R, http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf
[73] Hornik K and Hornik MK. Package 'NLP', 2018, http://cran.rstudio.org/web/packages/NLP/NLP.pdf
[74] Jovi A, Brki K and Bogunovi N. An overview of free software tools for general data mining. In: *37th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, Opatija, Croatia, 26–30 May 2015, pp.1112–1117. IEEE.
[75] Taylor R. Interpretaion of the correlation coefficient: a basic review. *J Diagn Med Sonogr* 1990; 6: 35–39.
[76] Bermudez-Edo M, Barnaghi P and Moessner K. Analysing real world data streams with spatio-temporal correlations: entropy vs. Pearson correlation. *Autom Constr* 2018; 88: 87–100.
[77] Ju A, Jeong SH and Chyi HI. Will social media save newspapers? *Journal Pract* 2014; 8: 1–17.