NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação

Master Program in Statistics and Information Management

Multivariate Data Analysis and social media: A Contribution to Infodemic Management Optimization Strategy

Susanah Bernardo da Silva Diniz

Dissertation presented as partial requirement for obtaining the master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Multivariate Data Analysis and social media: a Contribution to Infodemic Management Optimization Strategy

Susanah Bernardo da Silva Diniz

Proposal Dissertation presented as partial requirement for obtaining the master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Multivariate Data Analysis and social media: A Contribution to Infodemic Management Optimization Strategy

by

Susanah Diniz

Proposal presented as partial requirement for obtaining the master's degree in Statistics and Information Management, with a specialization in Information Management

Advisor: *Leonor Bacelar Nicolau*

November 2022

ABSTRACT

An infodemic is a huge flow of inaccurate and wrong information that may spread through social media, during an epidemic, potentially causing confusion and a damaging effect on peoples’ behavior and health. It also makes the intervention of public health agents more difficult. An infodemic can intensify outbreaks as it makes it hard for people to find trustworthy sources and reliable guidance when they need it.

The study's main objective was to characterize the individual engagement performance of social media posts published before and during the Covid19 pandemic (before and after vaccination) on Facebook’s pages of selected national health organizations in order to identify a typology of agencies.

Publicly available data on 39525 posts from 17 health agencies Facebook’s pages between 01/01/2019 and 31/05/2022 was retrieved and analysed with univariate and bivariate exploratory data analysis, text analysis methods and multivariate exploratory data analysis methods such as principal components analysis and hierarchical cluster analysis.

Results showed that globally the Covid19 pandemic led to a relevant increase in the number of posts published on the health agencies’ Facebook pages under study and also led to a large increase on the respective audiences’ interactions. However, there was a decrease in the engagement on the pandemic period after start of the vaccination, compared to the period of the actual pandemic. Furthermore, we identified 3 types of agencies: agencies with predominance performance in total interactions, agencies with higher and lower performance in relative engagement, and finally, agencies with an opposing performance between the pandemic period and the period of mass vaccination.

In short, with the Covid-19 pandemic, the public looked for more information through Facebook. Nonetheless, there might be a link between the differences in performance from these pages and different infodemics strategies.

Despite some limitations, our study provides valuable insights to health agencies and the public in general, as the infodemic management should not end after the crisis but should be an ongoing investment and may represent one of the best ways to make a more effective and competent health promotion.

KEYWORDS

Covid-19; Infodemiology; Infoveillance; Infodemic, health literacy.

List of Abbreviations and Acronyms

EPI-WIN	Information Network for Epidemics
NCD	non-communicable diseases
WHO	World Health Organization
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
COVID-19	Coronavirus Disease
PCA	Principal Component Analysis
CTR	Partial contribution of the variables or the individuals for a principal component
CO2	Squared cosines of each variable or individual with the principal components
KPI	Key Performance Indicators

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Study Relevance and Problem Identification

Since 2019 the world is fighting an epidemic of coronavirus disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) coronavirus. The fight against the disease impacted every aspect of our lives, from simple things such as limiting our holiday travels or meeting our friends to far-reaching consequences such as halting the global economy, causing millions of people to lose their jobs and income (Smith & Judd, 2020). In parallel to coronavirus, a virus of fake news and incorrect information is also spreading all over the world, for example seeding doubts in the public about the effectiveness of imposed measures or vaccination (Naeem et al., 2021). Anyone can be a spreader of fake news, by simply liking a friend’s post to sharing a news article. Unfortunately, fake news has a real-world impact, that resulted in lower vaccination takes in many countries. Such an epidemic is called an infodemic.

Infodemic is defined as an overabundance of misinformation which is rapidly propagated potentially causing confusion and a damaging effect on peoples’ behavior and health. The World Health Organization (WHO) has called it a disease accompanying the COVID-19 epidemic and has a massive impact on our daily lives and our ability to fight the health epidemic (Zarocostas, 2020). Furthermore, the infodemic may also promote fake cures, social panic, irrational fear, racism, xenophobia, and mistrust in the authorities among others.

Consequently, the infodemic intensifies outbreaks as it makes it hard for people to find trustworthy sources and reliable guidance when they need it (Tangcharoensathien et al., 2020).

One of the main problems the infodemic causes is decreasing the effectiveness of public safety measures put in place by public bodies to reduce the spread of COVID-19. The infodemics, i.e., the spread of misinformation, have a severe impact on public health (Tasnim et al., 2020). For example, the study by (Islam et al., 2020) showed that 5.800 people were admitted to hospital as a result of covid-19 misinformation disseminated on social media. Similarly, the study by (Bel Trew, 2020) highlighted that people were drinking methanol because they read on the internet that it could cure covid-19. Additionally, the spread of fake news also causes panic and intensifies xenophobia (Doj, 2020).

To summarise, misinformation prevents people from having effective health behaviors and people may decrease trust levels in healthcare professionals (Nsoesie & Oladeji, 2020). Therefore, the authorities must recognize the problem and employ various techniques to combat it, and be proactive in disseminating and promoting trusted and reliable information (Ding et al., 2020).

Consequently, health authorities are taking urgent actions to combat the spread of misinformation and promote infodemic management. For example, the World Health

Organization (WHO) established the Information Network for Epidemics (EPI-WIN) to act as a network connecting technical and social media teams within WHO (WHO Competency Framework Building a Response Workforce to Manage Infodemics, 2021). EPI-WIN disseminates and amplifies evidence-based information about COVID-19, and tracks and responds to misinformation, myths, and rumors.

Additionally, among many other responses to health promotion as courses and reports (Situation Report-51 SITUATION IN NUMBERS Total and New Cases in Last 24 Hours, 2022), WHO held an online technical consultation to receive suggestions concerning actions on responding to the infodemic related to covid-19 pandemic (Tangcharoensathien et al., 2020). However, it may be questionable if the national health authorities are indeed being efficient in their communication with the audiences, particularly on social media such as Facebook.

The impact of infodemics is mitigated by replacing misinformation with real, factual, information, or as the disinformation expert Claire Wardle from Harvard University in Cambridge, Massachusetts, stated, “The best way to fight misinformation is to swamp the landscape with accurate information that is easy to digest, engaging and easy to share on mobile devices” (Timothy Caulfield, 2020).

As we will show in more detail in the next chapter, further analysis is necessary to see if the health agencies had a satisfactory or poor performance on Facebook in propagating a sound message.

1.2 STUDY OBJECTIVES

1.2.1 Research Objectives

In this master thesis, we will characterize the individual engagement performance of social media posts published before and during the Covid19 pandemic on Facebook pages of selected national health organizations. We will monitor the main publicly available key performance indicators (KPIs) such as total interaction, likes, comments, shares, relative engagement given the number of page followers at the posting date, etc., between 01/01/2019 until 31/05/2022 in order to identify a typology of posts. As a secondary objective, we will compare the evolution between the specified periods and agencies to identify engagement profiles.

1.2.2 Research questions

1. Did the social media Facebook approach of international public health agencies changed before and during the COVID-19 pandemic?

2. Do people interact and engage on Facebook differently before and during the COVID-19 pandemic?

1.2.3 Research Hypothesis

The following main hypotheses are under study:

Hypothesis 1: Publishing strategies are different between the health institutions over the 4 periods under study.

We will associate the periods (P1: Pre-pandemic - 01 /01/2019 to 31/12/2019, P2: Before the declaration of pandemic – 01/01/2020 to 10/03/2020, P3: Before vaccination -11/03/2020 to 07/12/2020, and P4: during mass vaccination – 08/12/2020 to 31/05/2022) with the quantitative analysis of the volume of interactions. We expect that more posts are published on Facebook during the pandemic compared to before, which may be due to different publishing strategies of health agencies.

Hypothesis 2: The engagement of posts during pandemic is higher than before pandemic. Yet, there are differences between the health institutions, periods, and types of engagement (total interactions and relative engagement).

We wish to see if the pandemic might have changed the way the public interacts with national health agencies on Facebook. We expect to identify the agencies that may have been more, and less, successful regarding their health promotion Facebook communication approach.

Hypothesis 3: Publishing strategies and engagement rate are different with the start of vaccination.

We expect that, with the beginning of the vaccination, the health agencies publishing strategies may have changed and their audiences' engagement may have been different.

Hypothesis 4: Through the textual analysis of the posts' text, an evolution in the use of words "covid" and "vaccination" may be identified, regarding the frequency use and associated engagement, over the periods under study.

We expect to see a more frequent use of the words "covid" during the pandemic period (P2 and P3) and a more frequent use of words "vaccination" during the period of mass vaccination (P4).

In summary, our main goal is to identify a typology of Facebook's performance of national health organizations under study throughout the specified periods. More specifically we wish to check if the following types may be identified:

1. Health organizations that already had a good engagement with the public before covid-19 and continue to maintain their performance during the pandemic.
2. Health organizations that had a good engagement with the public before covid-19 but during the pandemic didn't manage to keep their performance.
3. Health organizations that didn't have any engagement with the public before covid-19 but during the pandemic managed to have a successful performance.
4. Health organizations that didn't have any engagement with the public before nor during covid-19.

The remainder of this thesis is structured as follows:

Chapter 2: **Background** Chapter presents main prior work and related literature on the topic of infodemic, also explaining how the infodemic can be characterized and managed.

Chapter 3: **Data and Methods** identifies the data collected and statistics methodologies applied in our research.

Chapter 4: **Results** chapter will display the results obtained through our data analysis.

Chapter 5: **Discussion** chapter will discuss the obtained results, put forward possible interpretations, and identify main limitations and potential paths for future research.

Chapter 6: **Conclusion** is the last chapter in which we summarise our findings.

2. BACKGROUND

In this chapter, we present the essential tools we rely on in our work. We provide a basic description of discussed topics and provide reference to numerous manuscripts in which a reader may find a more in-depth description of them.

This chapter consists of four sections. In the first section, we explain what Infodemic is and why it is important. Then a section on **Infodemiology** gives further detail on what is the science behind Infodemic. In the third section, we discuss **Health Literacy** which is a key to better health promotion. Finally, we conclude the chapter with a section on **the Infodemic Management Transdisciplinary model**. In the latter, we point out why this thesis contributes to the science community on this topic.

2.1 INFODEMIC

The year 2020 will be remembered primarily for the Covid 19 pandemic. While the pandemic is well known to society, the associated infodemic remains relatively unknown. In this section, we will thus provide a formal definition of infodemic and explain its meaning.

The term "infodemic" is composed of the words "information" and "epidemic" and was coined by Eysenbach in 2002 when the SARS outbreak shook the world (Eysenbach, 2020). Infodemic is defined as "an excess of misinformation, some of which is accurate and some of which is not, making it difficult for people to find trustworthy sources and reliable guidance when they need it” (Donovan, 2020). An infodemic can amplify outbreaks and may cause mass panic and fear.

Covid- 19 is an alarming problem, both in terms of its spread and severity (WHO, 2020). Given its spread, people must have access to trusted information to help them understand the crisis, protect themselves, and take appropriate action. It is clear that how people understand and respond to a public health crisis, and how they judge what actions are or aren’t appropriate, is shaped by information and all kinds of misinformation (Kleis Nielsen et al., 2020). As researchers have long known, it is risk perception, not actual risk, that determines how people respond to crises (Glik, 2007).

Such views are also reinforced by the quantity and quality of information people receive, which affects their knowledge and perception, which in turn influences their actions in terms of prevention and control (Geldsetzer, 2020). Another study (Ding et al., 2020) examines real-world examples of Covid-19 infodemics. The most common ones are SARS-CoV-2 virus was created as a biological weapon in a laboratory, that drinking bleach or eating garlic cures the infection, and that radio waves from 5G technology cause the sickness (Ella Hollowood & Alexi Mostrous, 2020).

Concerns about an infodemic or outbreak of misinformation predate Covid-19. For example, the authors (Tran & Lee, 2016) used Twitter to collect posts about the Ebola virus and found that 58.9% of the posts were misinformation. Another study (Oyeyemi et al., 2014) examined the spread of the Ebola infodemic and found that misinformation was more prevalent on social media than correct information.

Therefore, combating an infodemic became a matter of utmost importance whenever an epidemic outbreaks. The scale and importance of the problem led international organizations such as the World Health Organization and the United Nations to call it the first global infodemic in February 2020. "We're not just fighting an epidemic, we're fighting an infodemic"(Zarocostas, 2020).

In summary, the evidence on infodemics points us to the need for public health authorities to inform the public of trusted sources and reliable guidance when they need it.

2.2 INFODEMIOLOGY

Fighting an infodemic may be a complex task to undertake. You cannot eliminate an infodemic, although you may try to control it. With social media and the rapid spread of information, managing an infodemic becomes even more difficult.

“The proposal to fight the infodemic by spreading “facts” is easier said than done when it is not clear what the exact facts are.” (Eysenbach, 2020)

Infodemiology is a combination of the words “information” and “epidemiology” and is defined as “the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the aim to inform public health and public policy" (Eysenbach, 2009). The term was coined in 2002 by Eysenbach in the American Journal of Medicine, and it was considered a newly emerging research discipline and methodology to manage infodemic.

Originally, the concept of infodemiology aimed to identify the gap between expert knowledge and public practice (Eysenbach, 2002), it has since evolved to identify and analyse health information on the web through publicity shared searches, blogs, websites, and social media posts (Purnat et al., 2021).

In addition, infodemiology can be useful in guiding health professionals and patients to quality health information on the Internet. Gunther said in the American Journal of Medicine that "Information epidemiology or infodemiology identifies areas where there is a knowledge translation gap between best evidence (what some experts know) and practice (what most people do or believe), as well as markers of "high-quality" information" (Eysenbach, 2020).

Examples of infodemiology applications include as defined in (Eysenbach, 2009) are: "the analysis of queries from Internet search engines to predict disease outbreaks; identifying and monitoring of public health relevant publications on the Internet (e.g., anti- vaccination sites, but also news articles or expert- curated outbreak reports); automated tools to measure information diffusion and knowledge translation; and tracking the effectiveness of health marketing campaigns".

2.3 INFOVEILLANCE AND SOCIAL LISTENING

The term "social listening" is sometimes used as a synonym for Infoveillance, but has also been more narrowly defined as "the process of identifying and assessing what is being said about a company, product, brand, or individual, within forms of electronic interactive media"(Anderson et al., 2017).

“Infoveillance/social listening has been identified as one of the pillars to fight an infodemic” (Tangcharoensathien et al., 2020)

Currently, most emergency and outbreak recommendations emphasize the value of listening to communities, engaging them early in the response, and communicating clearly with them promptly (Risk Communication and Community Engagement Readiness and Response to Coronavirus Disease (COVID-19), 2020). Health agencies, therefore, face the challenge not only of providing relevant, high-quality health information but also of providing it at the right time, in the right format, and with community involvement (Purnat Tina, 2020).

Social listening can help overcome barriers to accepting quality health information and implementing healthy behaviors by increasing understanding of community issues, confusion, and information seeking, or by increasing awareness of specific issues (Purnat et al., 2021).

Social listening, as the name implies, is the process of following and analyzing social media platforms to gain insights and discover opportunities for action. Social listening has been used for applications where infodemiology methods are used for surveillance purposes and it is important to understand the public's questions, concerns, and misinformation. In addition, health professionals may also need to be aware of information demand inundation, whether to combat "epidemics of fear"(Eysenbach, 2003) by providing appropriate information to the public.

2.4 Health Literacy

“Health literacy is a term to describe the degree to which individuals can obtain, process, and understand basic information to make appropriate health decisions” (Liu et al., 2020). It is especially typical among older adults, minority populations, medically underserved people, and low socioeconomic class communities. For instance, according to (Paakkari & Okan, 2020) health literacy is globally important to prevent non-communicable diseases (NCD) and it is important to emphasize the need for people to be more responsible in managing their health by using effective use of health services.

Although the term first being proposed in 1970 (Simonds, 1974), over the past two decades health literacy has become increasingly important due to the significant benefits to the individual and public health and the reliability of the health care systems. It helps the public to understand and navigate a complex health system minimizing the risk of miscommunication.

Therefore, “health literacy is a relatively new concept in health promotion” (Don Nutbeam, 2006). The term health promotion may be defined as “the process of enabling people to increase control over and to improve their health” (Health Promotion Education & Communication, 1998). Health promotion supports governments, communities, and individuals to cope with and address health challenges. This is accomplished by building healthy public policies, creating supportive environments, and strengthening community action and personal skills (Van den Broucke, 2021).

With the COVID-19 pandemic, there was an overwhelming impact on society creating the perception that the existing healthcare systems are failing to protect communities against the spread of the virus. Hence, more people need to regain control of their health and abide by the health authorities to enhance protective behavior amongst citizens (Van den Broucke, 2021).

Thus, the importance of trust in dealing with crisis situations is the key to making better health promotion. For example, how African countries responded to the Ebola epidemic, shows that trusting environment helps to improve the understanding of the disease protocols (Marais et al., 2016).

In conclusion, improving people's access to health information and their capacity to use it effectively by making health care accessible to all, regardless of individual ability, is a fundamental requirement for empowerment. Moreover, this is one of the most important goals of effective and competent health promotion.

2.5 The Infodemic Management Transdisciplinary model

In this final topic of interest for this study we introduce and describe the Transdisciplinary Model of Infodemic Management.

The model was developed by the World Health Organization and is structured to incorporate the view of health authorities. This contrasts with other models that are built from a community, scientific development, or systems-level perspective (WHO Competency Framework Building a Response Workforce to Manage Infodemics, 2021). By placing the perspective of health authorities at the center of the model, it becomes important to articulate how science and evidence can inform health authority processes to improve adherence to public health and social care interventions, and the role of infodemic managers in these processes.

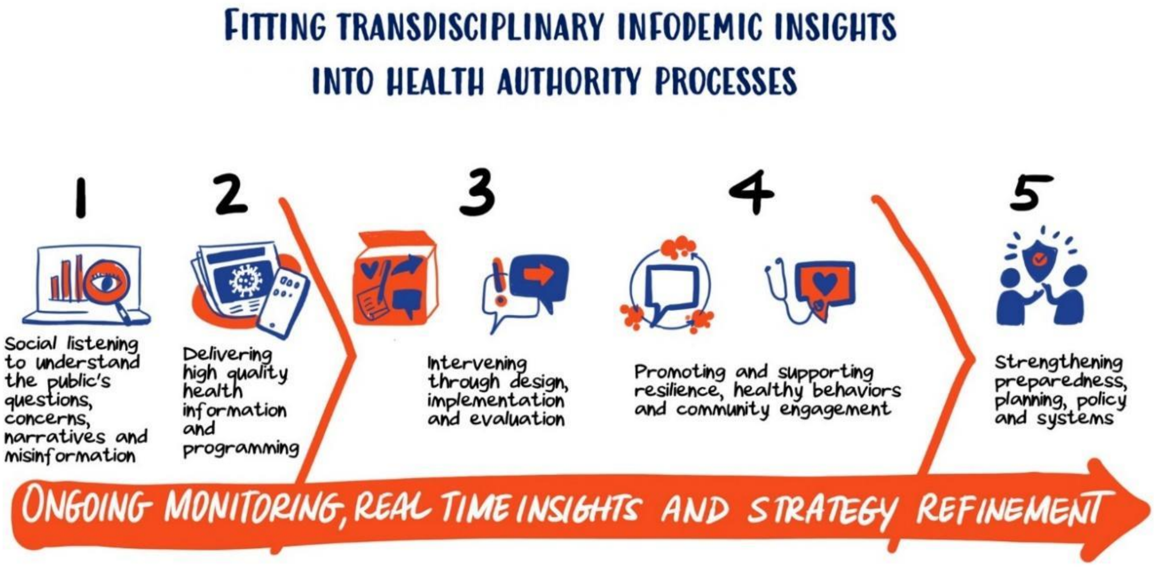


Figure 1 - Infodemic Management Model (WHO Competency Framework Building a Response Workforce to Manage Infodemics, 2021)

The framework is conceptualized around the five workstreams for infodemic preparedness and response along an epidemic curve, analogous to an epidemic response (Rubinelli et al., 2022). It is structured to benefit all personnel working in health institutions and organizations addressing the infodemic.

The objective of the model is to orient and support the institutions to strengthen Infodemic management capacity. It is a reference tool for activities such as linking work functions to required competencies and conducting training needs assessments and developing training plans (Calleja et al., 2021) .

The model has 5 steps:

Table 1- Steps of Infodemic Management Model

Social Listening	Listen, identify, and understand population gaps, needs, behaviors, and their determinants to develop more responsive health programs.
Delivering high quality health information and programming	Proactively share real, believable, and relevant information with target audiences to increase understanding, to build and strengthen health literacy, and to promote healthy behaviors on health issues.
Intervening through design implementation and evaluation	Offer corrections in a timely way that matches how the mis/disinformation is spread.
Promoting and supporting resilience, healthy behaviors, and community engagement	Support individuals' and communities' resilience against mis/disinformation.
Strengthening preparedness, planning, policy, and systems	Ensure that data-based insights and lessons learned from interventions are applied to prepare health systems with planning, processes, and policies for Information management.

As our study is based on the performance of the health agencies to engage their audiences, our research will focus mostly on steps that are more relevant for it and may mainly generate insights for steps 2, 3, and 4.

In step 2**, delivering high-quality health information and programming,** which is proactively sharing relevant information with target audiences to increase understanding and strengthen health literacy, will generate insights for the analysis of what is published on the Facebook pages of the health agencies.

In step 3**, intervening through design implementation and evaluation,** which is checking facts and trends over time and building or strengthening reporting tools and processes to identify and analyze misinformation, will generate insights for us to make a critical evaluation of the health agencies' publications, despite of being unaware of the effective institutions' strategies.

Finally, in step 4, **promoting and supporting resilience, healthy behaviors, and community engagement,** which is measuring community and evaluating interventions to build resilience against misinformation tailored to individual communities and vulnerable populations, will generate insights for the analysis of the engagement, in which we want to see if health agencies audiences' engagement may have been different than before the pandemic.

Furthermore, a growing body of literature on social media platforms has been used to address the infodemic. Social media data derived from Facebook, Instagram, Twitter, etc. provides useful information to pinpoint context-specific issues in real-time to allow for the fast identification of public attitudes(Moon & Lee, 2020). However, there is no evidence of studies that do a deeper analysis of the health agencies' attitudes and strategies toward their audience. Thus, our investigation makes remarkable contributions to the existing literature.

In this chapter, we have described, defined, and discussed essential terms for the context of our research, such as Infodemics, Infodemiology, Infoveillance/Social Listening, and Transdisciplinary Model of Infodemics Management. We will now analyze the data of interest using the methodologies described in the next section.

3. DATA AND METHODS

3.1. Data Sources

In this study, we collected data from seventeen national health agencies (Anvisa, Ministerio da Saúde, Direção Geral de Saúde, Serviço Nacional de Saúde, Instituto Nacional de Saúde Doutor Ricardo Jorge, NIJZ - Nacionalni inštitut za javno zdravje, Ministerio de Sanidad, Santé publique France, Institut National de la Santé et de la Recherche Médicale, Ministère des solidarités et de la Santé, Helse- og omsorgsdepartementet, Helsedirektoratet, Sundhedsstyrelsen, Socialstyrelsen, and Karolinskainstitutet) from 9 countries (Brazil, Portugal, France, Italy, Spain, Slovenia, Denmark, Sweden, and Norway) between 01/01/2019 and 31/05/2022, as shown in Table 2.

Data were retrieved through CrowdTangle, a public insights tool powered by Facebook that tracks data on public pages. CrowdTangle tracks "interactions on public content from Facebook pages and does not include paid ads unless those ads began as organic, unpaid posts that were subsequently "boosted" using Facebook's advertising tools. It also does not include activity on private accounts or posts that are only visible to certain groups of followers." (https://help.crowdtangle.com/en/articles/3192685-citing-crowdtangle-data)

Key Performance Indicators (KPIs) or metrics data were retrieved on 01/06/2022 for an analysis period from 01/01/2019 to 31/05/2022, which includes 39.550 posts. We defined 4 periods from 2019 to 2022 considering the most important dates during the Covid 19 pandemic:

- **Period 1:** from 01 /01/2019 to 31/12/2019 which is the period before the COVID -19 pandemic.
- **Period 2:** from 01/01/2020 to 10/03/2020, it is the period between when WHO reported a cluster of pneumonia cases in Wuhan, and when the COVID -19 pandemic WHO's declaration.
- **Period 3**: from 11/03/2020 to 07 /12/2020, it is the period of the pandemic before the start of the vaccination (08/12/2020 in UK)
- **Period 4**: from 08/12/2020 to 31/05/2022, it is the period of the pandemic during mass vaccination.

Variables describing each post include Page name, date the post was published, time period the post was published, followers at the time of publication, total interactions, relative engagement (for 100.000 page likes at the time of publication), likes, comments, shares, message, mention of "covid" in the message, and mention of "vaccination" in the message. Further information on KPI and content variables may be found in Table 3.

Relative engagement (for 100.000 page likes at the time of posting) is determined by dividing total interaction by the number of "page likes" at the time of posting times 100.000. Because the Facebook pages under study have widely varying numbers of "page likes" and have varying levels of interaction, the relative engagement analysis is particularly important. It provides insight into how well the pages are succeeding in appealing to their "natural" audience, i.e., those who have liked the page, may follow the posted content more regularly, and are more likely to interact with the published posts.

Table 2 - List of Health agencies under study

Country	Agency name
Brazil	Anvisa
	Ministerio da Saude
Portugal	DGS
	Serviço Nacional de Saúde
	Instituto Nacional de Saúde Doutor Ricardo Jorge
Slovenia	NIJZ - Nacionalni inštitut za javno zdravje
Spain	Ministerio de Sanidad
Italy	Istituto Superiore di Sanità
	Ministero della Salute
France	Santé publique France
	Institut National de la Santé et de la Recherche Médicale (Inserm)
	Ministère des solidarités et de la santé
Norway	Helse- og omsorgsdepartementet (Norge)
	Helsedirektoratet
Denmark	Sundhedsstyrelsen
Sweden	Socialstyrelsen
	Karolinskainstitutet

For the multivariate exploratory data analysis (3.2.3), we used a dataset obtained from the previously described dataset, focusing only pandemic period variables (mean of the variables during the pandemic period - Total interactions, Likes, Shares, Comments and Relative Engagement for periods 3 and 4), and as cases the health agencies under study.

Metrics/KPIs and variables describing each post include:

Variable	Definition
Page Name	Name of the agency that public a post on Facebook.
Post Created Date	Date which the post was created.
Page Admin Top Country	In each country the post was published.
Period when Post was created	P1: January 2019 – December 2019 **- Period pre-covid**
	P2: January 4 to March 10, 2020 - **Period before WHO declaration of COVID-19 Pandemic**
	P3: March 11, 2020, to December 7, 2020 - **Period before vaccination**
	P4: December 8, 2020, to May 31, 2022 **- Period during mass vaccination**
Total Interactions	The sum of the total number of reactions (likes, love, wow, haha, sad, angry, and care emojis), comments and shares for each post, during the selected time range.
Message	Text posted on the post. It does not include the text of the comments, only the main original text posted.
Mention "Covid" in message	Mention of "Covid" in message: identification if the post included the word stem "COVID" on its main message or not.
Mention "Vaccination" in message	Mention of "Vaccination" in message: identification if the post included the word stem "Vaccine" on its main message or not.

Comments	Number of instances where a user reacts to a post with a comment, including comment replies.
Likes	Number of instances where a user reacts to a post with a like. Other reactions, such as heart, sad, angry, haha, wow, or care emojis, are not included here.
Shares	Number of Instances when a user clicks “share” on a post.
Followers at Posting	Total of page followers on the date each post was published.
Relative Engagement	Total Interactions by the number of Likes at posting times 100.000. This KPI identifies how engaging the post is.

Table 3 - List of KPI's used for this thesis

3.2. Statistical Analysis

3.2.1. Univariate and bivariate exploratory data analysis

We performed univariate and bivariate exploratory data analysis (means, standard deviations, medians, and interquartile ranges for quantitative variables and frequencies and percentages for qualitative variables), building tables and graphical representations to facilitate the interpretation of results. We analyzed the post-publication, interaction, and relative engagement by posting date, page, time period, and mentions of the word stem "covid" and "vaccination" in the message.

3.2.2. Text analysis methods

Additionally, we applied text analysis methods to extract the main word stems and phrases from the text published for each post. The latter allowed us to create word clouds to look for insights for each time period under study.

In creating the word clouds, the size of each word was proportional to the number of mentions of the word in posts published during each observed time period. The color grading of each word was related to the relative engagement rate in the posts where the word was mentioned. The grading scale was determined for each time period considering the mean, standard deviation, maximum and minimum for that time period.

3.2.3. Multivariate Analysis

Multivariate exploratory data analysis methods were used, namely, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis, to identify typologies and groups of agencies with similar performance over the pandemic periods (before and during vaccination).

3.2.3.1. PCA

PCA is a powerful data analysis tool, capable of reducing large complex data sets containing many variables to a few principal components set that allows us to spot main underlying trends and patterns.

The selected active variables used for the PCA were: Mean of: Relative Engagement, Total interactions, comments, shares and, likes for period 3 and 4 (pandemic period).

As the measurement scales, as well as the means and standard deviations of these variables, are not similar, the PCA was based on the correlation matrix.

To choose the components for the analysis and the main criteria used were:

1. **Pearson Criterion:** the components should be selected until the cumulative proportion is at least 80% of total inertia.
2. **Scree Plot Criterion:** the components should be selected until the eigenvalues stabilize (left of the 'elbow' on Figure 9).
3. **Kaiser Criterion: since** the PCA of this work is based on the correlation matrix, the components with eigenvalues greater than 1 should be selected.
4. **Squared Cosines (CO^2) coefficients for variables and cases:** When choosing the principal components for interpretation, we assured the sums of the CO^2 for each variable and case over the selected axes were always higher than 80%, to guarantee the variables and cases were were sufficiently explained with the number of selected axes (Table 9).

To interpret each of the selected principal components, the most relevant variables were identified through the CTR (partial contributions of variables) of the variables, which showed how much each variable contributes for the total inertia of each axis. The most relevant variables and cases to explain a given axis were determined by if their CTRs are over the average CTR (100/10 = 10) (Table 10). Moreover the squared cosines of the variables, is considered a quality indicator, because it shows how much inertia of each variable is explained by each component. Thus, we identified aditional relevant variables as the ones with at least 50% of inertia in each principal component.

3.2.3.2. Cluster analysis

Finally, a cluster analysis was applied, with the objective of making a typology of the health agencies. We used the dataset of the coordinates of the previously selected principal components obtained through the PCA outputs. We applied an ascending hierarchical cluster analysis, using the Euclidean distances as proximity measures and the Ward method as aggregation criteria. The groups for analysis were chosen with the help of the hierarchical tree, the evolution table of the distances in each aggregation, the table with the evolution of the CCs, and the remaining outputs of the methodology.

The software used for the analysis was JMP®, version Pro 16.0.0. Citation: JMP®, version Pro 16.0.0. SAS Institute Inc, Cary, NC, 1989-2021.

In the next chapter, we show the results of the data analysis performed using the methods described in this chapter.

4. RESULTS

In this chapter, we present the results of our data analysis. In each of the four sections, we focused on a different aspect of public interaction with the seventeen Facebook pages of selected national health organizations across the four time periods. In the first section, we highlighted publication and interaction KPIs. This is followed by a section in which we examined relative interaction KPIs. Since we analyzed pages with widely varying numbers of likes and interactions in this section, the analysis of relative engagement was particularly important. In the third section, we focused on text analysis and identify posts with the words "covid" and "vaccination". Finally, we performed a multivariate exploratory analysis based on Principal Component and Hierarchical cluster analysis in the fourth and last section of this chapter.

4.1. Post publications and interactions

The total number of posts published throughout the 17 Facebook pages under study during the four periods explored is shown in Table 5. The period before pandemic (P1) accounts for 19% of the total 39,525 posts published (7,314 posts). However, the pandemic period (P2, P3 and P4) accounts for 81% of the total 39,525 posts published (32,211 posts).

During period 3 (pre-vaccination pandemic) there was an increase of 1.4 times more posts published (10,315 posts, an average of 606 posts published in period 3) compared to P1 (7,314 posts, average of 430 posts published). During P4, after the introduction of vaccination, the posts published continued to exhibit a higher performance (20,143 posts, average of 1,184 posts published in period 4), there was an increase of 2 and 2.75 times more compared to period P3 and P1, respectively (Table 4).

This discrepancy between the number of posts published and the extent of each interaction in each period is very clear in Table 4 and 5. P1 account for 19% of the total 39,525 posts published in the four time periods, but only 3% of the total 77,288,302 interactions. P2 accounts for 4% of the total 39,525 posts published in the four periods and 2% of the total 77,288,302 interactions. Conversely, P3 accounts for 26% of the total 39,525 posts published in the four time periods, and 68% of the total 77,288,302 interactions. Period 4 accounts for 51% of the total 39,525 posts published in the four periods but only 28% of the total 77,288,302 interactions.

A total of 7,314 posts published on these 17 Facebook pages on P1, before the pandemic, resulted in 1,981,930 interactions, an average of 2,901 interactions per post. This average nearly tripled over the course of P2 (January 01-March 10, 2020), increasing to 1,235,488 interactions before WHO confirmed the pandemic outbreak. With the declaration of the COVID-19 pandemic, the average interaction per post increased fifteenfold (44,847) during P3 compared to P1 to 52,207,994 interactions. After vaccination began, the average interaction per post decreased slightly but remained high at 13,894 interactions per post, 4.8 times higher than during the first period (Table 7, Figure 2).

The performance of each of the 17 pages under study was not consistent across the different time periods. Before the pandemic outbreak, i.e., period P1, the Ministerio da Saúde had an average of 815 interactions (reactions, comments, and shares), followed by Socialstyrelsen with an average of 293 interactions, the Ministero della Salute and the DGS with an average of 287 and 234 interactions, respectively.

During P2, before WHO declared the pandemic, Ministero della Salute and DGS had an average of 1,676 and 1,218 interactions, respectively, six times more than during P1. On the other hand, Socialstyrelsen had a 0.5-fold decrease in average interactions (193).

During P3, the most critical phase of the pandemic, Ministerio de Sanidad, Ministère de la santé and Ministerio da Saúde had an average of 5,456, 4,583, and 21,457 post interactions, respectively, 41, 49, and 26 times more than in P1. Socialstyrelsen and NIJZ, on the other hand, saw a decrease in average interactions to 0.60 and 0.76 of the average interactions in P2, respectively.

After the start of vaccination, in P4, there was an overall decrease in average interactions to 0.30 of the average interactions in P3. However, compared with P1, there was an increase in overall average interactions after vaccination, with 13,894 average interactions in P4, 4.78 times more than the average interactions in P1 (2,901). For example, Ministère de la santé and DGS, with 13 and 10 times more average interactions after vaccination, respectively.

Ministerio da Saúde, Ministère de la santé, and Ministero della Salute presented the best interaction performance, which was to be expected given the larger number of followers on their page. On the other hand, NIJZ, Karolinska Institutet and Helsedirektoratet showed the worst interaction performance.

The analysis of relative engagement in the next section could provide further insight into whether this performance could also be associated with possible more effective digital strategies.

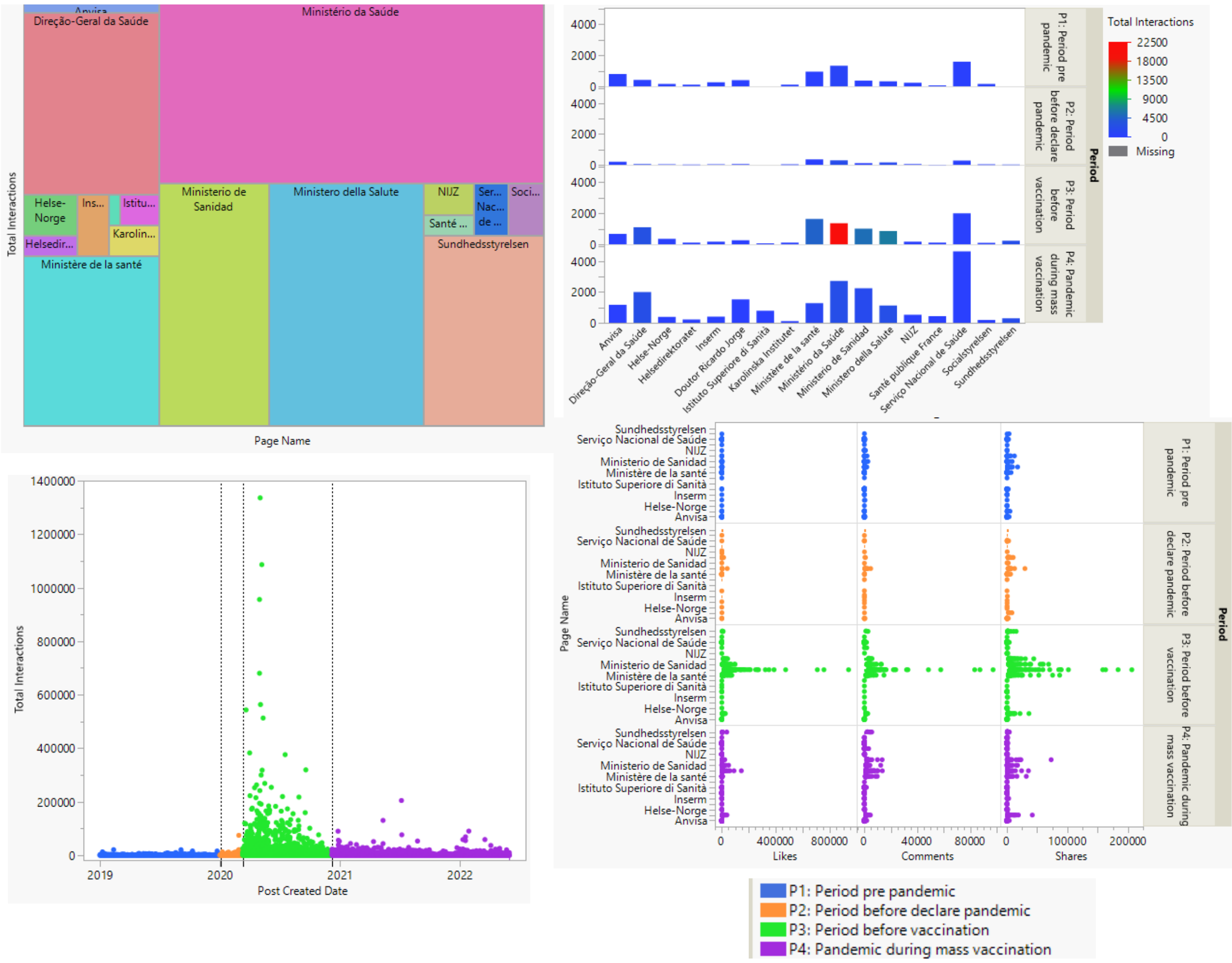


Figure 2 – Total interactions by Page name, Total interactions by period and post created date, Likes, Comments and Shares by period and page name.

Table 4 - Number of posts published and total interactions per period

	Total Interactions				
	N	Sum	Mean of posts published per period	Percentage of total posts	Percentage of total interactions
P1: Period pre pandemic	7,314	1,981,930	430	19%	3%
P2: Period before declare pandemic	1,753	1,235,488	103	4%	2%
P3: Period before vaccination	10,315	52,207,994	**606**	26%	68%
P4: Pandemic during mass vaccination	20,143	21,862,890	**1,184**	51%	28%
Total	**39,525**	**77,288,302**			

Table 5 - Number of posts published per page name by period

	P1: Period pre pandemic	P2: Period before declare pandemic	P3: Period before vaccination	P4: Pandemic during mass vaccination
Anvisa	802	215	681	1,176
Direção-Geral da Saúde	420	61	1,096	2,014
Helse-Norge	160	31	357	390
Helsedirektoratet	119	17	108	221
Inserm	272	34	178	407
Doutor Ricardo Jorge	403	57	262	1,529
Istituto Superiore di Sanità	0	0	61	795
Karolinska Institutet	115	30	108	109
Ministère de la santé	952	355	1626	1,285
Ministério da Saúde	1,335	294	1363	2,724
Ministerio de Sanidad	368	116	1,009	2,257
Ministero della Salute	321	154	852	1,134
NIJZ	234	59	174	529
Santé publique France	68	8	119	436
Serviço Nacional de Saúde	1,590	283	1,990	4,652
Socialstyrelsen	155	27	102	192
Sundhedsstyrelsen	0	12	229	293

Table 6 - Number of followers per page per period

	P1: Period pre pandemic	P2: Period before declare pandemic	P3: Period before vaccination	P4: Pandemic during mass vaccination
	Followers at Posting			
	Mean			
Anvisa	86,324	97,661	114,415	153,046
Direção-Geral da Saúde	100,647	109,976	523,878	766,315
Helse-Norge	26,561	28,862	39,207	54,408
Helsedirektoratet	48,935	51,315	98,750	114,321
Inserm	27,592	31,162	37,456	46,996
Doutor Ricardo Jorge			18,255	19,630
Istituto di Sanità			2,147	23,936
Karolinska Institutet	35,159	37,059	38,797	42,669
Ministère de la santé	76,596	90,593	1,043,375	1,422,187
Ministério da Saúde	2,164,390	2,204,860	4,864,874	5,376,670
Ministerio de Sanidad	92,555	101,594	683,468	933,206
Ministero della Salute	112,120	195,202	1,133,914	1,506,232
NIJZ	33,443	41,151	74,612	90,266
Santé publique France	11,103	15,084	32,776	46,911
Serviço Nacional de Saúde	163,545	177,405	232,587	273,627
Socialstyrelsen	14,845	17,194	19,358	22,248
Sundhedsstyrelsen		2,480	97,562	171,096

Table 7 - Total interactions per period per page name

			DGS	SNS	Doutor Ricardo Jorge	Ministério da Saúde	Anvisa	Ministerio de Sanidad	Ministero della Salute	Istituto di Sanità	Ministère de la santé	Santé publique France	Inserm	Helsedirek toratet	Helse-(Norge)	NIJZ	Karolinska Institutet	Socialstyr elsen	Sundhedssty relsen
P1	Total Interactions	Mean	234	189	51	815	110	131	287		92	118	110	111	107	124	129	293	
		Median	110	89	28	476	59	21	87		45	68	84	47	52	64	79	169	
		Std Dev	489	335	207	1,184	263	1,211	1,087		293	137	107	161	203	209	153	557	
		Q1	46	37	16	263	31	12	44		25	41	47	21	28	31	45	80	
		Q3	231	200	51	911	111	35	189		88	134	134	144	103	120	164	358	
		Sum	98,470	301,242	20,620	1,087,670	88,390	48,324	92,004		87,862	8,000	29,802	13,168	17,079	28,975	14,849	45,475	
		N	420	1,590	403	1,335	802	368	321		952	68	272	119	160	234	115	155	
P2	Total Interactions	Mean	1,218	312	56	2,100	210	163	1,676		234	88	199	125	83	249	120	193	962
		Median	409	127	40	1,051	95	60	664		86	75	135	64	55	136	100	84	339
		Std Dev	1,814	519	65	4,831	346	347	2,804		704	44	182	158	115	378	78	232	1,257
		Q1	197	50	23	541	47	33	263		42	52	77	33	33	63	54	47	110
		Q3	1,460	346	60	2,263	239	129	1,855		195	128	326	157	80	316	190	211	1,799
		Sum	74,320	88,289	3,184	617,355	45,089	18,868	258,157		83,025	700	6,760	2,118	2,574	14,708	3,588	5,215	11,538
		N	61	283	57	294	215	116	154		355	8	34	17	31	59	30	27	12
P3	Total Interactions	Mean	2,891	258	22	21,457	118	5,456	6,416	58	4,583	154	245	156	293	190	154	117	2,279
		Median	1,649	139	14	3,451	59	2,247	3,642	24	1,196	67	158	59	175	94	94	67	1,137
		Std Dev	3,695	357	28	72,182	198	13,886	8,369	133	11,192	460	279	267	453	298	216	144	3,556
		Q1	762	77	11	1,314	34	869	1,616	16	564	43	91	36	86	48	58	42	502
		Q3	3,851	317	22	12,603	116	5,055	8,032	52	3,917	108	299	176	330	194	167	129	2,377
		Sum	3,168,545	514,026	5,695	29,246,270	80,363	5,504,921	5,466,449	3,545	7,451,390	18,354	43,548	16,840	104,471	33,055	16,685	11,908	521,929
		N	1,096	1,990	262	1,363	681	1,009	852	61	1,626	119	178	108	357	174	108	102	229
P4	Total Interactions	Mean	2,515	124	32	2,416	140	1,750	2,545	118	1,216	75	212	72	190	148	150	103	2,088
		Median	1,322	66	24	892	57	785	1,950	60	578	32	141	25	138	60	84	58	1,346
		Std Dev	3,517	155	34	6,271	324	3,357	4,586	191	2,359	359	306	161	264	266	235	146	3,528
		Q1	498	35	14	351	31	427	908	33	270	21	77	13	64	31	49	35	665
		Q3	3,406	153	37	2,428	132	1,793	2,677	120	1,157	56	254	56	228	140	169	112	2,484
		Sum	5,065,218	576,020	48,310	6,581,422	164,781	3,950,037	2,886,071	93,455	1,562,003	32,831	86,246	16,018	74,283	78,304	16,391	19,752	611,748
		N	2,014	4,652	1,529	2,724	1,176	2,257	1,134	795	1,285	436	407	221	390	529	109	192	293

4.2. Relative Engagement

The total average performance throughout the 17 Facebook pages under study during the four periods explored is shown in Table 7.

The period before pandemic (P1) accounts for 17% of the total average performance (6,517 number of posts by 100,000 page likes of relative engagement per post, an average of 383 number of posts by 100,000 pages likes). P2 and P3 accounts for 4% and 27% of the total average performance (1,691 and 10,307 interactions by 100,000 page likes of relative engagement per post, an average of 99 and 606 number of posts by 100,000 pages likes). However, period 4, accounts for 52% of the total average performance (20,143 interactions by 100,000 page likes of relative engagement per post, an average of 1,184 number of posts by 100,000 pages likes).

A total of 6,517 number of posts by 100,000 page likes of relative engagement per post on these 17 Facebook pages on P1, before the pandemic, resulted in 1,512,814 interactions. This got 8 times more (11,595,535 interactions by 100,000 page likes of relative engagement per post) over the course of P3 (11/03/2020 – 07/12/2020), with the declaration of the COVID-19 pandemic. After vaccination began, the interactions by 100,000 page likes of relative engagement per post decreased slightly but remained high at 5,585,098 interactions by 100,000 page likes of relative engagement per post, 3.7 times higher than during the first period (Tables 8, 9 and Figure 3).

On absolute terms, NIJZ, Karolinska Institutet and Helsedirektoratet registered lower publishing and interaction, and Ministerio da Saúde, Ministère de la santé, and Ministero della Salute presented higher publishing and interaction. However, the relative engagement complements these observations with additional insights.

During P1, Socialstyrelsen presented the highest average performance (2,101 interactions by 100,000 pages likes of relative engagement per post), on the other hand, Ministerio da Saúde presented the lowest average performance (38 interactions by 100,000 pages likes of relative engagement per post).

During P2, when the existence of an epidemic in Wuhan was acknowledged but the pandemic was not yet declared, most of the agencies improved their average engagement performance, especially Sundhedsstyrelsen, with 49,048 interactions by 100,000 pages likes per post. However, Socialstyrelsen encountered a decrease in average engagement performance to 0.56 of the average engagement in P1.

With the declaration of the COVID-19 pandemic, we expected a changed on the previous engagement settings completely. For instance, Ministerio da Saúde, Ministero de Sanidad and Inserm had an increase of 8.1, 12.8 and 6.8 times more in average engagement performance than in Period 2. However, NIJZ and Socialstyrelsen had a decrease in average engagement performance to 0.46 and 0.55 of the average engagement in P2. Although, Sundhedsstyrelsen had a decrease in average engagement performance to 0.15 of the average engagement in

P2, this agency performance was still higher (7,336 interactions by 100,000 pages likes per post) compared to the other agencies as Istituto Superiore di Sanità with 3,893 interactions by 100,000 pages likes per post.

After the start of the vaccination, the average overall engagement decreased to 0.33 of the overall average engagement in P3. And again, the agency Sundhedsstyrelsen had a decrease to 0.3 of the overall average engagement in P3 thus continued to have a higher engagement performance (2,488 interactions by 100,000 pages likes per post) compared to the other agencies.

After the analysis of relative engagement, we could conclude that the agencies that had a good total interactions performance, didn’t necessarily have a good relative engagement one. For example, Ministerio da Saúde, which was a top rated in total interactions, had the worst performance in relative terms. Sundhedsstyrelsen, on the other hand, had the best relative performance and not very well in total interactions. This will be discussed further in chapter 5 in the discussion.

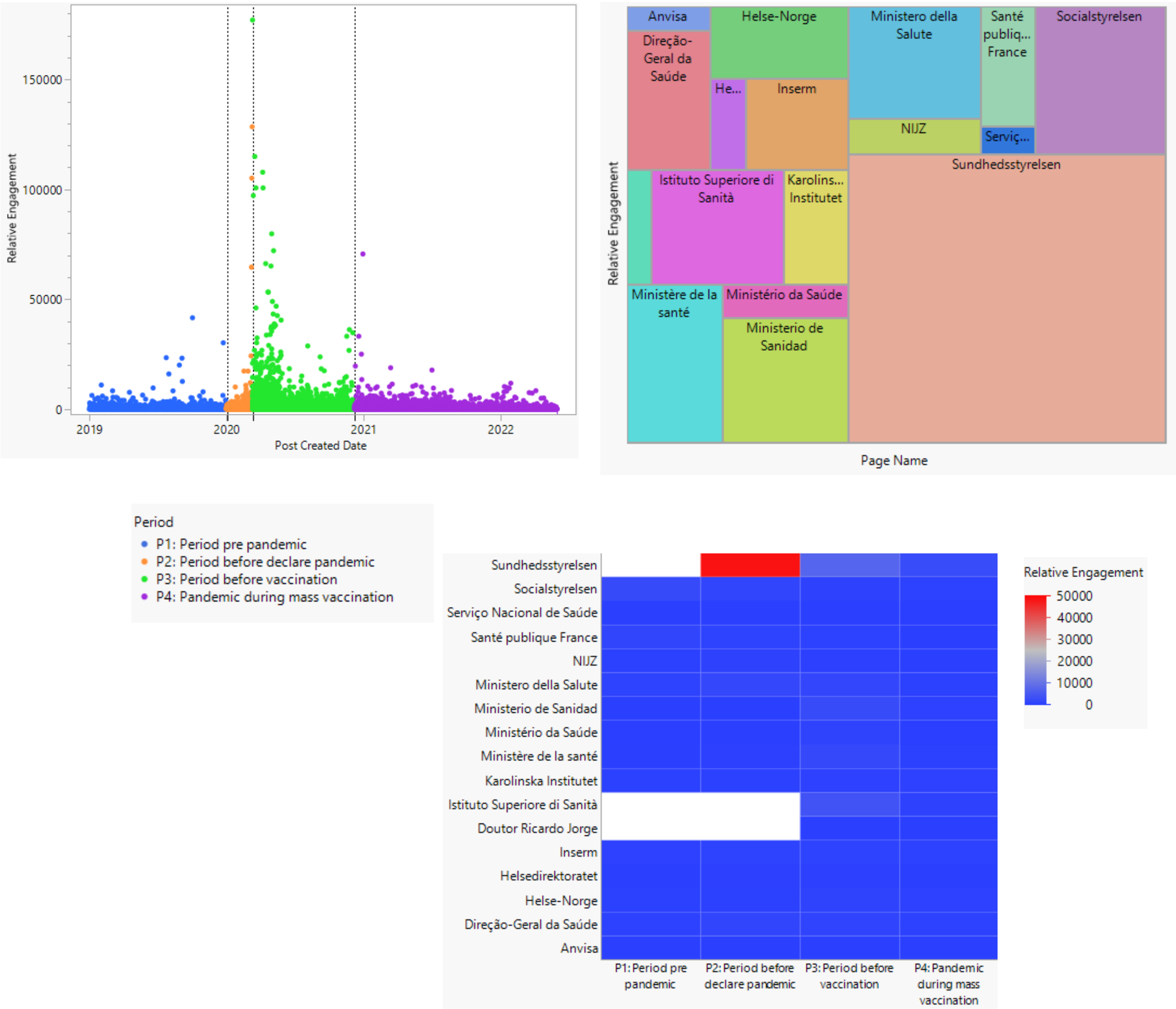


Figure 3 – Relative Engagement by Page name, Relative Engagement by post created date and Relative Engagement by period and page name.

Table 8 - Number of posts by 100.000 pages likes per post and total interactions by period

	Relative Engagement			
	Number of posts by 100.000 pages likes	Number of interactions	Mean of posts by 100.000 pages likes per period	Percentage of total average performance
P1: Period pre pandemic	6,517	1,512,814	383	17%
P2: Period before declare pandemic	1,691	1,026,642	99	4%
P3: Period before vaccination	10,307	11,595,536	606	27%
P4: Pandemic during mass vaccination	20,143	5,585,099	1,184	52%
Total	38,658	19,720,090		

Table 9 - Relative Engagement per page name by period

			DGS	SNS	Doutor Ricardo Jorge	Ministério da Saúde	Anvisa	Ministerio de Sanidad	Ministero della Salute	Istituto Superiore di Sanità	Inserm	Ministère de la santé	Santé publique France	Helsedirektoratet	Helse-Norge	NIJZ	Karolinska Institutet	Socialstyrelsen	Sundhedsstyrelsen
P1	Relative Engagement	Mean	528	117		38	133	150	716		127	1,241	421	234	416	375	381	2,101	
		Median	206	56		22	71	23	239		61	676	314	99	203	194	231	1,141	
		Std Dev	835	207		55	323	1,379	2,424		412	1,406	412	345	802	619	450	3,893	
		Q1	126	23		12	37	14	113		34	369	175	44	110	92	135	567	
		Q3	497	121		42	131	40	498		119	1,641	509	297	400	367	482	2,447	
		Sum	13,721	185,474		50,546	106,932	55,016	229,731		120,741	84,357	114,577	27,875	66,549	87,751	43,831	325,714	
		N	26	1,590		1,335	802	368	321		952	68	272	119	160	234	115	155	
P2	Relative Engagement	Mean	1,122	178		96	224	178	1,615		276	683	687	253	300	605	352	1,178	49,048
		Median	410	73		48	100	67	679		106	590	468	131	197	355	295	509	24,267
		Std Dev	1,700	296		222	371	386	2,623		813	345	625	323	422	925	231	1,417	51,068
		Q1	190	29		25	49	36	295		52	417	268	64	118	158	157	288	7,428
		Q3	1,357	196		104	250	143	1,751		228	977	1,110	319	287	741	559	1,301	105,156
		Sum	68,440	50,446		28,362	48,087	20,665	248,669		98,112	5,463	23,346	4,303	9,308	35,719	10,567	31,817	343,338
		N	61	283		294	215	116	154		355	8	34	17	31	59	30	27	7
P3	Relative Engagement	Mean	937	124	133	783	111	2,277	1,445	3,893	1,867	542	727	275	844	283	455	647	7,336
		Median	579	67	77	126	55	816	747	1,593	441	239	464	100	506	141	277	368	2,710
		Std Dev	1,127	176	195	2,659	198	7,085	2,307	6,788	5,024	1,430	873	468	1,320	444	637	801	16,504
		Q1	255	36	61	45	31	316	352	688	192	143	262	62	243	72	171	232	1,435
		Q3	1,236	150	133	478	106	1,920	1,636	3,422	1,447	426	890	324	954	279	491	714	6,598
		Sum	1,026,647	246,210	34,647	1,067,543	75,897	2,297,740	1,230,764	210,222	3,036,530	64,555	129,327	29,656	301,438	49,254	49,145	65,998	1,679,964
		N	1,096	1,990	261	1,363	681	1,009	852	54	1,626	119	178	108	357	174	108	102	229
P4	Relative Engagement	Mean	508	52	172	81	102	423	355	668	296	196	504	114	425	180	439	508	2,488
		Median	262	27	132	30	40	183	265	304	141	84	340	39	304	74	245	290	1,568
		Std Dev	747	65	184	210	236	866	626	1,618	572	948	757	253	617	325	689	706	4,649
		Q1	94	14	77	12	21	97	127	150	66	53	179	21	137	37	143	176	772
		Q3	676	63	204	81	94	428	378	631	277	149	608	88	497	170	493	545	2,986
		Sum	1,023,976	239,661	262,827	220,126	119,705	954,182	402,102	530,950	380,280	85,467	205,210	25,174	165,839	95,375	47,899	97,443	728,884
		N	2,014	4,652	1,529	2,724	1,176	2,257	1,134	795	1,285	436	407	221	390	529	109	192	293

4.3. Text Analysis

After comparing periods and agencies regarding posts publishing, total interactions and relative engagement performances, a more qualitative analysis was undertaken, focusing the text of the published posts. For each period, word clouds were built and mentions of “covid” and “vaccination” were counted. Note that the largest words found reflect the greater or lesser number of posts published by certain agencies.

The word cloud (Figure 4) obtained for the pre-pandemic period P1 is very diversified. The 10 largest stem words more often mentioned in the posts (largest in the word cloud), such as “saúde”, “nacional”, “gov”, “anvisa”, “sns”, are not the ones with a higher engagement (colours grading more often to grey and blue tones). Naturally, no mention of “covid” was found on posts during P1.

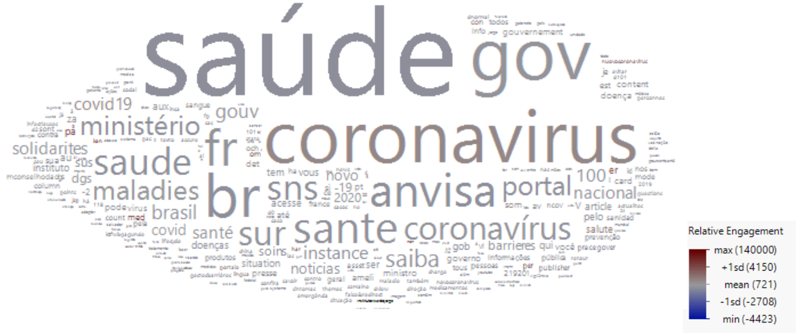
The words “saúde” and “gov” are the ones that are more often mentioned during P2 (after the WHO acknowledgement of the existence of an epidemic in Wuhan, but before the Pandemic declaration), but present an engagement below the period’s average (Figure 4). The other more mentioned stem words, such as “coronavirus” have generated a higher engagement, presenting a reddish color.

During P3 (covering the pandemic period before vaccination started, from 11/03/2020 to 07/12/2020), the word that really stands out in the word cloud is “covid”, “covid19”, “saúde”, with an engagement slightly higher than the mean (Figure 4). The word “coronavirus” has generated a higher engagement, presenting a reddish color.

With the beginning of the vaccination process (period P4), the obtained word cloud is very similar to the one on the previous period P3 (Figure 4). The words that stand out the most in the word cloud are still “saúde”, “Covid” and “covid19”, followed by word stem “vacinação”, but with smaller interaction.

P1 – Period pre pandemic

P2 – Period before declaration of the pandemic



P3 – Period pandemic before vaccination

P4 – Period pandemic during mass vaccination

Figure 4 - Word Clouds from text of posts published, over the 4 periods under analysis

Size of each stem word proportional to the number of times it was mentioned in posts published over each time period. Colour of each stem word graded according to the relative engagement observed for the posts where the word was mentioned. Grading scale determined for each period, considering the mean and standard deviation, maximum and minimum on each period.

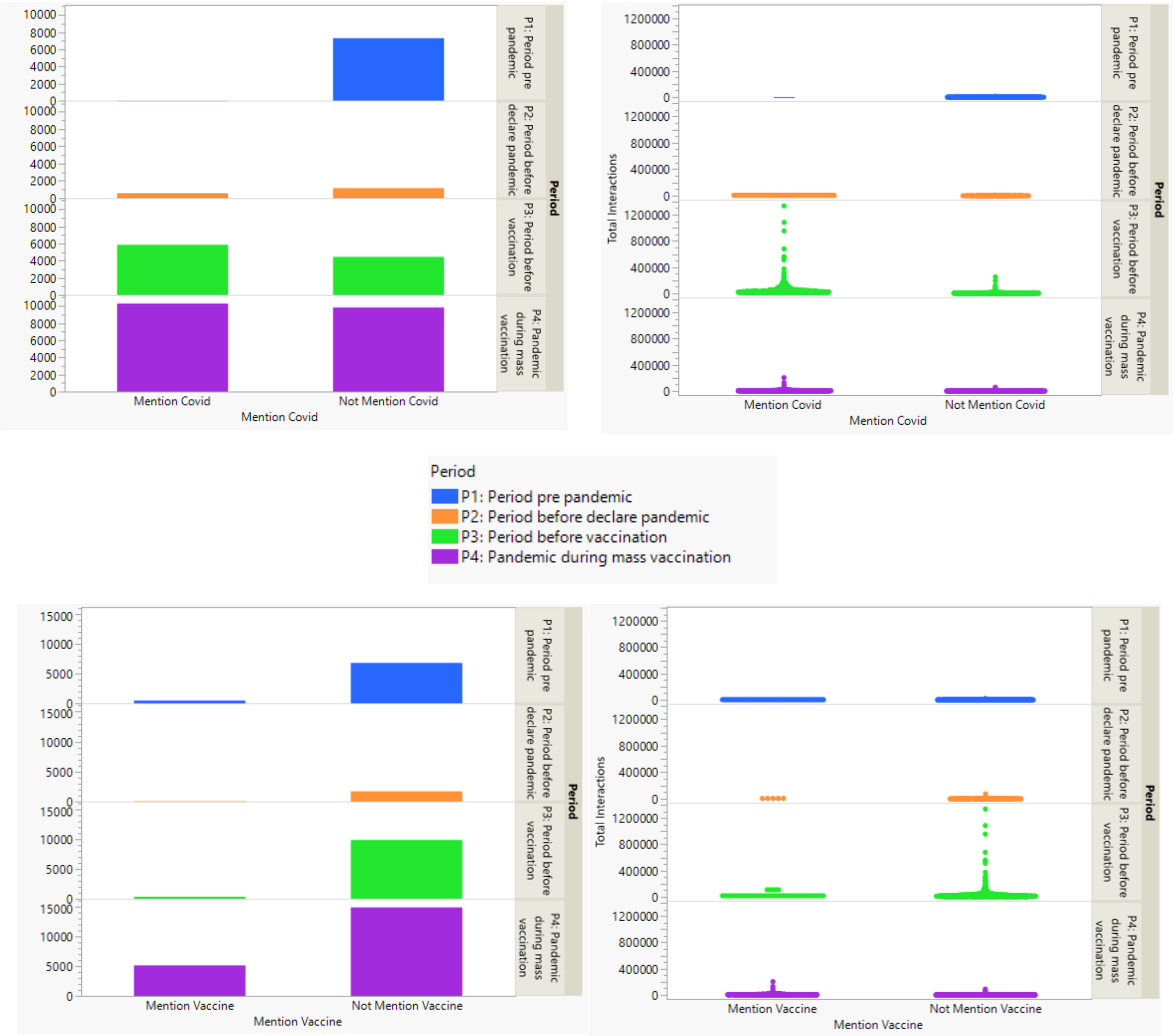


Figure 5 - Number of posts mentioning "covid" and "vaccination" and respective interaction, by Period

4.4. Multivariate Analysis

The univariate exploration preceding the multivariate analysis on data is important to identify possible asymmetries in the distribution of variables, that might create results biases.

We used a different dataset since we wanted to identify typologies of agencies, and not typologies of posts. Hence the cases are not posts, but agencies. Additionally, we wanted to focus on the effect of the pandemic, and thus only used variables regarding periods 3 and 4 (most critical pandemic periods).

The Boxplot analysis (Figure 6) provides a visualization of the empirical distribution of each variable, showing very assimetric, non-normal distributions of the variables. However, in this study, we will not perform any transformation on variables (Box-cox for example), to correct assimetries or treat outliers and we will use data as it is. In fact, the nature of these social media data makes these assimetries important, since outliers constitute essentially viral posts with higher engagement (Jansen et al., 2021) or in the case of this dataset agencies publishing posts that gone viral.

The PCA was based on the correlation matrix because the measurement scales, as well as the means and standard deviations of these variables are not similar.The Linear Correlation Matrix for the active variables is showned on Table 10.

The Bartlett's test of sphericity was used to test if there is a high correlation between the observed and expected values. Hence, as we can see in Table 11, we can assume that, in this case, the use of PCA is applicable. As the variables are highly correlated, PCA can be used to reduce the original variables into a smaller number of new variables (principal components) explaining most of the variance of the original variables.

The PCA outputs (Table 11) show that the variance of the first principal component is $\lambda 1=69479$, which means that the first principal axis of inertia accounts for 69.5%. On the other hand, the second principal component has a variance of $\lambda 2=2.0952$ and therefore the second principal axis explains 21% of the total inertia and the third principal component has a variance of $\lambda 2=0.6471$ and therefore the third principal axis explains 6.5% of the total inertia. These three first principal axes of inertia thus explain 96.9% of the total inertia and hence we consider it relevant to retain an additional third axis.

From Pearson and Kaiser criteria the first 2 components should be chosen for the analysis (Table 11). From the Scree plot (Figure 7) the first 3 components should be chosen for the analysis. The Squared Cosines (CO^2) coefficients indicate that all the variables are well explained by the 3 components, so we decided to use 3 principal components for our analysis (Table 13).

Table 16 presents the most relevant variables and individuals for the interpretation of each of the components. This can also be seen in the representation of the individuals/agencies on the first and second axes factorial plane and first and third axes factorial plane (Figure 9) and in the assotiated correlations circles (Figure 10).

The 1st Principal Component is a size factor, positively correlated with all variables, in which the most relevant variables are total interactions P3 and P4, comments P3 and P4, likes P3 and P4 and shares P3. An opposition is created between Ministerio da saúde, Ministère des solidarités et de la santé and Ministero della Salute with tendentially higher values for these variables and NIJZ, Helsedirektoratet and Karolinska Institutet with lower values.

The second principal component differentiates the relative engagement variables from the rest performance indicators, as Sundhedsstyrelsen and Socialstyrelsen display the best relative engagement behavior.

The third principal component contrasts the variables of P3 with those of P4, as Ministério da Saúde displays a better performance in period 3 and Ministero della Salute a better performance in period 4 and shows those that had similar performance in both periods as Karolinska Institutet and Santé publique France .

The hierarchical cluster analysis (Figure 10, Tables 17 and 18) allowed to identify a partition of 2 clusters (given the dendrogram and the evolution of distances for each partition step): one integrating less similar agencies (botton one in the dendrogram or hierarchical tree, aggregating later in the clustering process) and the other one subdivided into 2 more similar subgroups more cohese (aggregating earlier in the clustering process). We thus selected for analysis 3 clusters (Figure 10), then characterized them using bivariate statistical analysis.

Table 19 shows the main descriptive statistics of the clusters. The first cluster (red) includes the set of 6 agencies (Anvisa, Serviço Nacional de Saúde, Helsedirektoratet, NIJZ, Santé publique France and Instituto Nacional de Saúde Doutor Ricardo Jorge) with the lowest mean of engagement.

The second cluster (blue), with 5 agencies (Helse- og omsorgsdepartementet (Norge), Inserm, Istituto Superiore di Sanità, Karolinska Institutet, Socialstyrelsen) with an inclination for high engagement. Cluster three (green) includes 6 agencies (Direção-Geral da Saúde, Ministère des solidarités et de la Santé, Ministerio de Sanidad, Ministério da Saúde, Ministero della Salute and Sundhedsstyrelsen) with a higher mean of engagement.

Using PCA and HCH, complementary, we may see in Figure 13 that the PC1 distinguish the agencies from cluster 3 from the other agencies. PC2 opposes agencies with higher and lower engagement performance. PC2 and PC3 are relevant to distinguish cluster 1 from cluster 2.

Therefore, after presenting the key aspects of all the three cluster, we identify 3 types of agencies:

1. The agencies with higher and lower performance in total interactions.
2. The agencies with higher and lower performance in relative engagement.
3. The agencies with an opposing performance between period 3 (pandemic) and period 4 (mass vaccination).

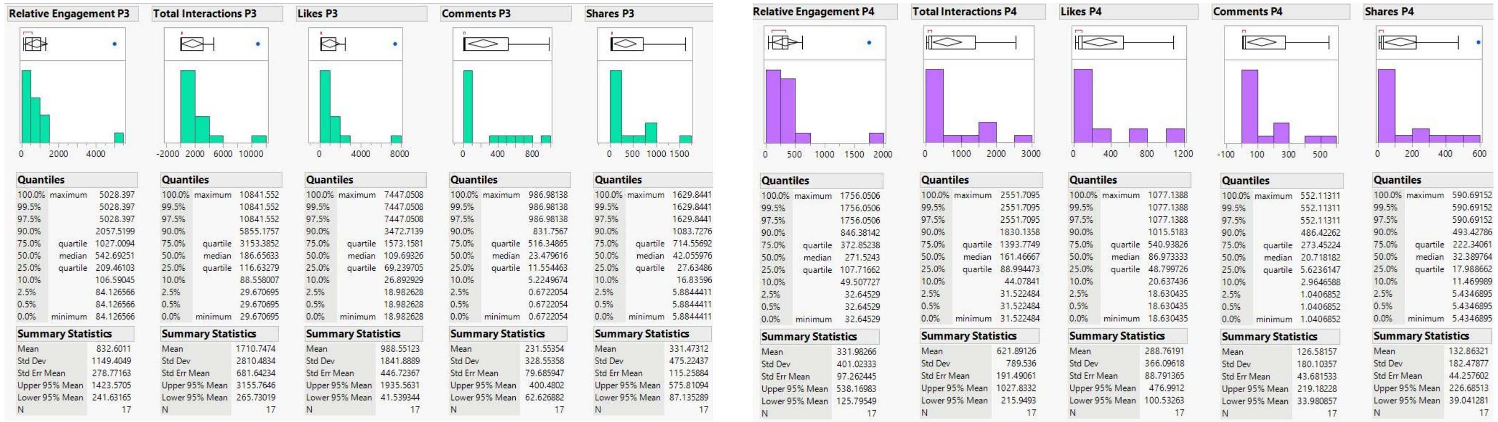


Figure 6 - Univariate analysis (boxplots, histograms, quantiles, and summary statistics) of the variables

Table 10 – PCA: Linear Correlation Matrix

	Relative Engagement P3	Total Interactions P3	Likes P3	Comments P3	Shares P3	Relative Engagement P4	Total Interactions P4	Likes P4	Comments P4	Shares P4
Relative Engagement P3	1	0.1403	0.0757	0.3994	0.1769	**0.9185**	0.472	0.4066	**0.6281**	0.3291
Total Interactions P3	0.1403	1	**0.9898**	**0.919**	**0.9744**	-0.0424	**0.748**	**0.8265**	**0.6743**	**0.5875**
Likes P3	0.0757	**0.9898**	1	**0.8612**	**0.9351**	-0.0806	**0.6743**	**0.7813**	**0.5845**	**0.5008**
Comments P3	0.3994	**0.919**	**0.8612**	1	**0.9509**	0.1832	**0.8876**	**0.8916**	**0.8972**	**0.7272**
Shares P3	0.1769	**0.9744**	**0.9351**	**0.9509**	1	-0.0392	**0.8135**	**0.8399**	**0.7516**	**0.7084**
Relative Engagement P4	**0.9185**	-0.0424	-0.0806	0.1832	-0.0392	1	0.3176	0.2799	0.4485	0.1796
Total Interactions P4	0.472	**0.748**	**0.6743**	**0.8876**	**0.8135**	0.3176	1	**0.9657**	**0.947**	**0.9202**
Likes P4	0.4066	**0.8265**	**0.7813**	**0.8916**	**0.8399**	0.2799	**0.9657**	1	**0.8779**	**0.817**
Comments P4	**0.6281**	**0.6743**	**0.5845**	**0.8972**	**0.7516**	0.4485	**0.947**	**0.8779**	1	**0.8332**
Shares P4	0.3291	**0.5875**	**0.5008**	**0.7272**	**0.7084**	0.1796	**0.9202**	**0.817**	**0.8332**	1

Table 11- PCA eigenvalues

Number	Eigenvalue	Percent	Cum Percent	ChiSquare	DF	Prob>ChiSq
1	**6.948**	69.479	69.479	517.815	39.464	<.0001*
2	**2.095**	20.952	90.430	396.089	42.032	<.0001*
3	**0.647**	6.471	96.901	299.458	35.568	<.0001*
4	0.168	1.679	98.580	225.004	28.260	<.0001*
5	0.094	0.940	99.520	186.500	20.902	<.0001*
6	0.036	0.362	99.883	144.500	14.852	<.0001*
7	0.011	0.113	99.996	104.261	9.219	<.0001*
8	0.000	0.004	100.000	32.864	5.182	<.0001*

Table 12 -PCA: Loading Matrix

	Prin1	Prin2	Prin3
Relative Engagement P3	0.451	0.848	0.213
Total Interactions P3	0.890	-0.373	0.257
Likes P3	0.832	-0.427	0.330
Comments P3	0.973	-0.085	0.113
Shares P3	0.926	-0.328	0.102
Relative Engagement P4	0.267	0.919	0.220
Total Interactions P4	0.960	0.117	-0.239
Likes P4	0.959	0.020	-0.076
Comments P4	0.925	0.279	-0.137
Shares P4	0.830	0.067	-0.523

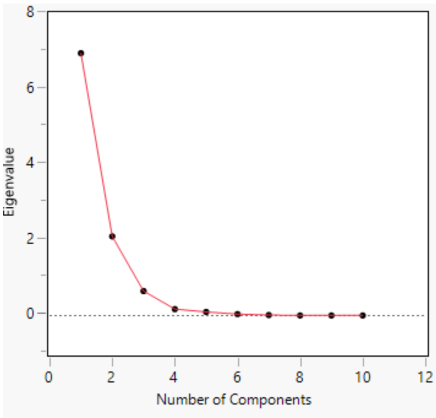


Figure 7 - PCA: Scree plot

Table 13- PCA Squared cosinus of the variables

	Prin1	Prin2	Prin3
Relative Engagement P3	0.203	**0.718**	0.045
Total Interactions P3	**0.792**	0.139	0.066
Likes P3	**0.693**	0.182	0.109
Comments P3	**0.946**	0.007	0.013
Shares P3	**0.858**	0.108	0.010
Relative Engagement P4	0.071	**0.844**	0.049
Total Interactions P4	**0.921**	0.014	0.057
Likes P4	**0.920**	0.000	0.006
Comments P4	**0.855**	0.078	0.019
Shares P4	**0.689**	0.005	0.274

Table 14 – PCA: CTR of the variables

	Prin1	Prin2	Prin3
Relative Engagement P3	2.927	**34.291**	7.002
Total Interactions P3	**11.401**	6.630	**10.241**
Likes P3	**9.968**	8.690	16.788
Comments P3	**13.614**	0.348	1.966
Shares P3	**12.346**	5.140	1.616
Relative Engagement P4	1.026	**40.290**	7.502
Total Interactions P4	**13.251**	0.657	8.810
Likes P4	**13.244**	0.020	0.887
Comments P4	**12.310**	3.718	2.901
Shares P4	**9.914**	0.217	**42.287**

Table 15 – PCA Outputs

Page Name	PC1	PC2	PC3	CO2 PC1	COS2 PC2	CO2 PC3	SUM OF CO2	CTR PC1	CTR PC2	CTR PC3
Anvisa	-1.843	-0.509	-0.142	**91.230**	6.962	0.541	98.733	3.056	0.773	0.195
Direção-Geral da Saúde	1.936	-0.267	-1.067	**65.680**	1.246	19.957	86.883	3.373	0.212	**11.004**
Helse- (Norge)	-1.554	0.246	0.188	**94.926**	2.380	1.392	98.699	2.172	0.181	0.342
Helsedirektoratet	-1.927	-0.438	-0.055	**94.417**	4.879	0.078	99.374	3.339	0.572	0.030
Inserm	-1.543	0.178	0.088	**96.672**	1.285	0.316	98.274	2.143	0.094	0.075
Instituto Doutor Ricardo Jorge	-2.014	-0.346	-0.001	**96.825**	2.864	0.000	99.689	3.648	0.358	0.000
Istituto Superiore di Sanità	-1.653	0.367	0.220	**89.894**	4.432	1.592	95.918	2.457	0.402	0.467
Karolinska Institutet	-1.629	0.298	0.191	**88.651**	2.966	1.216	92.833	2.386	0.265	0.351
Ministère des solidarités et de la santé	2.231	-0.073	-0.904	**70.686**	0.075	11.604	82.365	4.476	0.016	**7.890**
Ministério da Saúde	5.494	-2.967	1.863	**70.873**	20.677	8.149	99.699	**27.149**	**26.266**	**33.516**
Ministerio de Sanidad	1.482	-0.423	0.452	**54.535**	4.440	5.077	64.052	1.976	0.534	1.976
Ministero della Salute	4.998	-0.039	-1.870	**86.706**	0.005	12.138	98.849	**22.474**	0.005	**33.780**
NIJZ	-1.798	-0.374	-0.054	**94.515**	4.083	0.086	98.685	2.909	0.417	0.029
Santé publique France	-1.899	-0.360	-0.030	**95.695**	3.448	0.024	99.167	3.243	0.387	0.009
Serviço Nacional de Saúde	-1.908	-0.623	-0.108	**89.020**	9.492	0.286	98.799	3.275	1.158	0.113
Socialstyrelsen	-1.630	0.579	0.227	**76.875**	9.690	1.491	88.057	2.390	0.999	0.498
Sundhedsstyrelsen	3.256	4.752	1.003	30.982	**66.008**	2.943	99.934	**9.535**	**67.362**	**9.726**

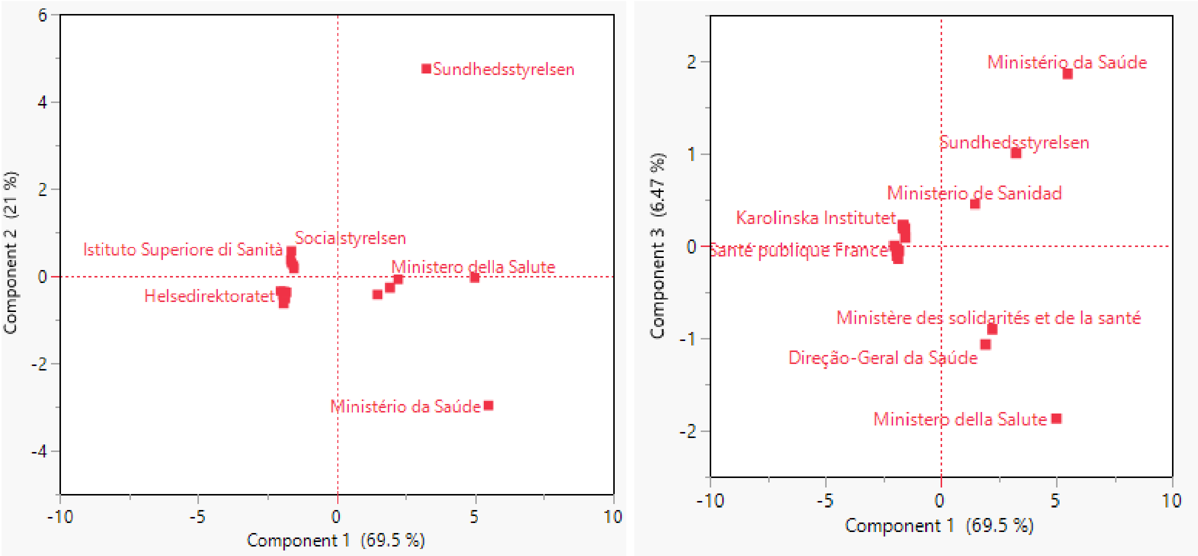


Figure 9 - PCA Representation of the individuals on the principal planes

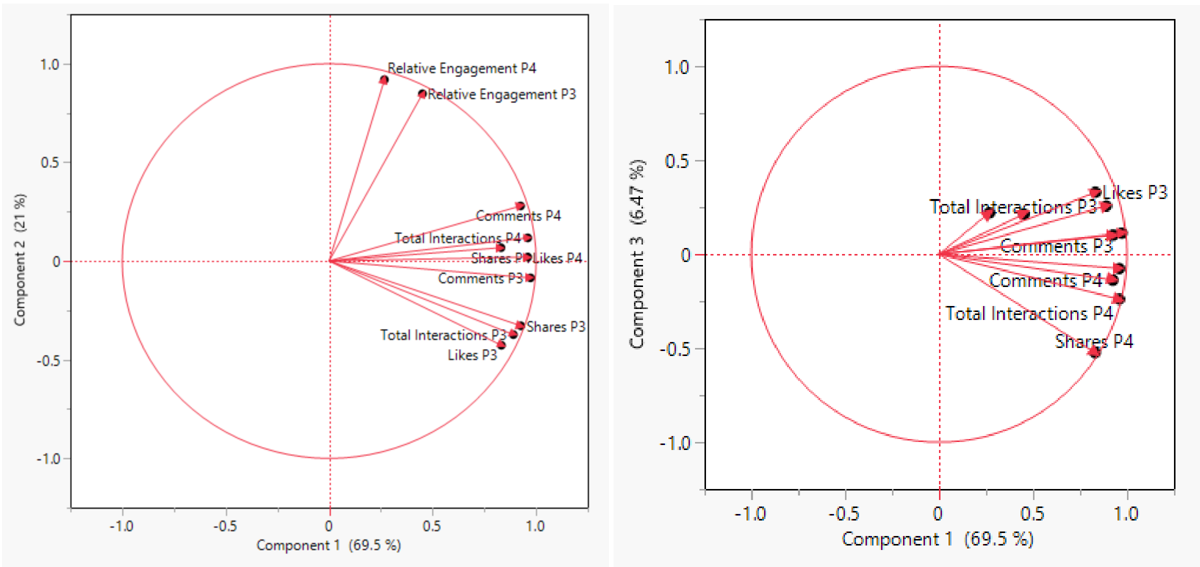


Figure 8 – PCA Correlation circles

Table 16 – PCA Contributions of Variables and Individuals per Principal Component

Component	Negative Contribution – CTR/COS2	Positive Contribution – CTR/COS2
PC1	Anvisa – 3.056% / 91.23%	Comments P4 -12.310%
	DGS – 3.373% / 65.68%	Total Interactions P4- 13.251%
	Helse- (Norge) – 2.172% / 94.926%	Likes P4- 13.244%
	Helsedirektoratet – 3.339% / 94.417%	Comments P3 - 13.614%
	Inserm – 2.143% / 96.672%	Shares P3 - 12.346%
	Instituto Doutor Ricardo Jorge – 3.648% / 96.825%	Total Interactions P3 - 11.401%
	Istituto Superiore di Sanità - 2.457% /89.894%	Likes P3 - 9.968%
	Karolinska Institutet –2.386% / 88.651%	Shares P4 – 9.914%
	NIJZ – 2.909% / 94.515%	Sundhedsstyrelsen – 9.535%
	Santé publique France – 3.243% / 95.695%	Ministero della Salute - 22.474%
	Socialstyrelsen – 2.39% / 76.875%	Ministère des solidarités et de la santé – 4.476% / 70.686%
	Serviço Nacional de Saúde – 3.275% / 89.02%	Ministerio da saúde - 27.149%
PC2		Relative Engagement P3 - 34.291%
	Ministério da Saúde – 26.266%	Relative Engagement P4 - 40.290%
	Serviço Nacional de Saúde – 1.158% / 9.492%	Socialstyrelsen – 0.999% / 9.69%
		Sundhedsstyrelsen - 67.362%
PC3	Shares P4 - 42.287%	Likes P3 - 16.788%
	Ministère des solidarités et de la santé - 7.89%	Total Interactions P3 – 10.241%
	Ministero della Salute- 33.78%	Ministério da Saúde - 33.516%
	Direção geral de saúde – 11.004%	Sundhedsstyrelsen – 9.726%

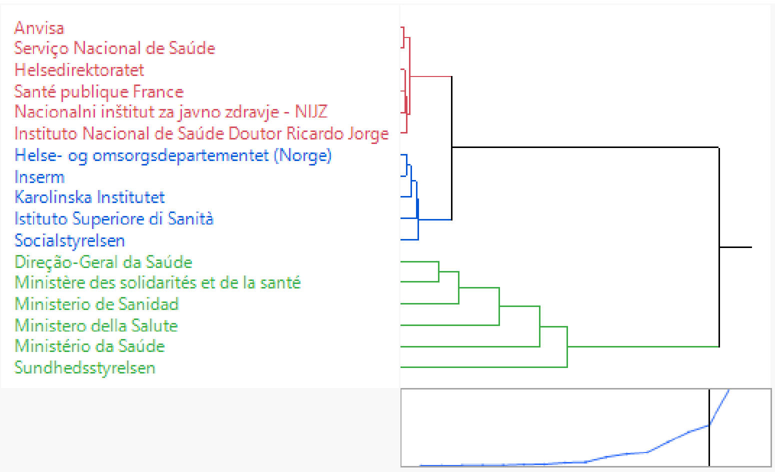


Figure 10 – Clustering Analysis - Dendogram of Hierarchical Cluster

Table 17 - Clustering Analysis - Evolution table of the Euclidean distances in each aggregation

Number of Clusters	Distance	Leader	Joiner
16	0.109	Anvisa	Serviço Nacional de Saúde
15	0.116	Helsedirektoratet	Santé publique France
14	0.139	Helsedirektoratet	Nacionalni inštitut za javno zdravje - NIJZ
13	0.162	Helse- og omsorgsdepartementet (Norge)	Inserm
12	0.170	Helsedirektoratet	Instituto Nacional de Saúde Doutor Ricardo Jorge
11	0.264	Anvisa	Helsedirektoratet
10	0.292	Helse- og omsorgsdepartementet (Norge)	Karolinska Institutet
9	0.482	Helse- og omsorgsdepartementet (Norge)	Istituto Superiore di Sanità
8	0.556	Helse- og omsorgsdepartementet (Norge)	Socialstyrelsen
7	1.177	Direção-Geral da Saúde	Ministère des solidarités et de la santé
6	1.569	Anvisa	Helse- og omsorgsdepartementet (Norge)
5	1.779	Direção-Geral da Saúde	Ministerio de Sanidad
4	3.067	Direção-Geral da Saúde	Ministero della Salute
3	**4.324**	Direção-Geral da Saúde	Ministério da Saúde
2	**5.172**	Direção-Geral da Saúde	Sundhedsstyrelsen
1	**9.866**	Anvisa	Direção-Geral da Saúde

Table 18 - Clustering Analysis - Evolution of CCCs

	Number of Clusters	CCC	
	1	0.000	
->	2	0.507	

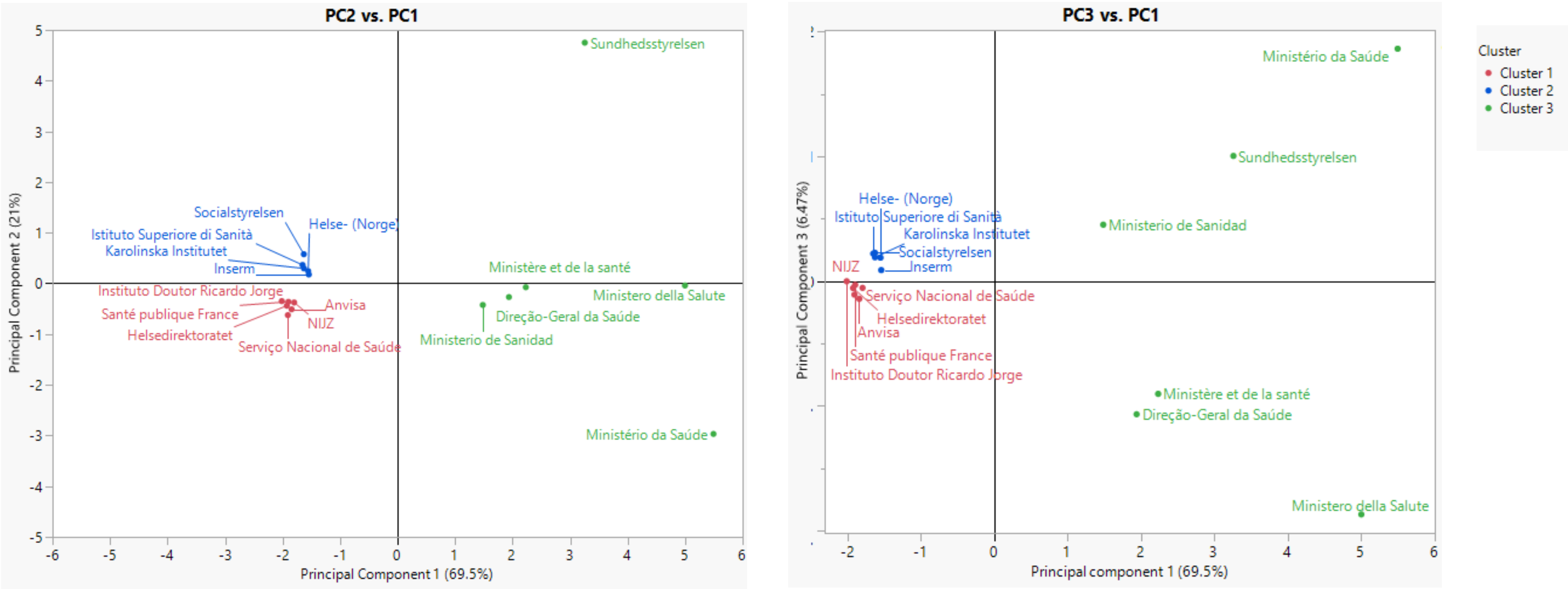


Figure 11 - Clustering Analysis -Representation of the clusters on the Principal Components Planes

Table 19 - Clustering Analysis - Main descriptive statistics of the clusters

		Cluster 1	Cluster 2	Cluster 3
Relative Engagement P3	Mean	184.474	679.945	1607.942
	Std Dev	85.297	257.077	1708.421
	Min	84.127	442.715	390.876
	Max	322.241	1112.902	5028.397
Total Interactions P3	Mean	124.404	172.791	4578.721
	Std Dev	56.191	67.994	3164.480
	Min	29.671	103.280	2247.355
	Max	186.656	247.077	10841.552
Likes P3	Mean	64.656	100.017	2652.891
	Std Dev	34.194	38.027	2390.821
	Min	18.983	62.365	1102.467
	Max	109.693	152.349	7447.051
Comments P3	Mean	19.641	13.835	624.899
	Std Dev	16.454	7.273	241.173
	Min	0.672	6.363	306.377
	Max	48.863	23.953	986.981
Shares P3	Mean	27.474	45.904	873.447
	Std Dev	13.826	26.857	420.694
	Min	5.884	26.580	387.715
	Max	44.516	92.414	1629.844
Relative Engagement P4	Mean	100.568	430.897	480.969
	Std Dev	43.125	121.504	632.668
	Min	32.645	311.007	53.723
	Max	157.234	618.964	1756.051
Total Interactions P4	Mean	73.012	146.217	1567.166
	Std Dev	30.765	33.135	576.981
	Min	31.522	97.071	801.748
	Max	109.061	179.488	2551.710
Likes P4	Mean	40.330	88.952	703.702
	Std Dev	16.683	31.214	327.461
	Min	18.630	52.339	288.299
	Max	58.606	127.630	1077.139
Comments P4	Mean	9.187	12.826	338.772
	Std Dev	9.700	6.522	141.997
	Min	1.041	7.392	174.335
	Max	22.524	23.907	552.113
Shares P4	Mean	15.380	33.336	333.286
	Std Dev	5.977	14.752	177.972
	Min	5.435	17.889	116.930
	Max	21.580	50.140	590.692

5. DISCUSSION

5.1. Research Hypothesis Discussion

We started this study with a research question aiming to find out if the social media Facebook approach of international public health agencies changed before and during the COVID-19 pandemic. The goal was achieved by fulfilling the four proposed research objectives.

The first hypothesis was focused on the publishing of posts and their interactions. We expected to see more posts published on Facebook during the pandemic compared to before. And what we found out was that globally, the Covid19 pandemic led to an increase in the number of posts published on the health agencies' Facebook pages under study, possibly as a result of more intense social media publishing strategies. This may be due to the realization of the importance of reliable social media information leading to an increase in posts published by these agencies in response to a higher interest of the general public seeking information on Covid19 (Bernardino & Bacelar Nicolau, 2020).

The second hypothesis was related to the engagement of the posts. Firstly, we expected to see a change in the way the public interacts with national health agencies on Facebook, during the pandemic. What we found out was that there was, indeed, a large increase of the total interactions during the Covid-19 pandemic. Secondly, when we analyzed the relative engagement performance, we encountered a scenario where the agencies who had a very high global performance (total interactions) didn't seem to engage so effectively with their audience. This could be due to the fact that the agencies didn't have the capacity to fully interact with their whole audiences and keep them engaged when these audiences are bigger and not so focused on the perhaps more specialized messages communicated (Alonso-Cañadas et al., 2020).

Hypothesis 3 concerned the start of the vaccination. We expected to see if, with the beginning of the vaccination, the health agencies' publishing strategies may have changed, and their audiences' engagement may have been different. What we found out was that there was a decrease in engagement, compared to the periods P3. This might be an indication that the audiences became less interested in the pandemic communication, especially with the end of restrictive and confinement measures (Centre for Disease Prevention, 2020), but also that agencies did not keep their engagement performance after the vaccination or might not have adjusted their message to a time where audiences should have been engaged to try and increase the vaccination rate (Ecdc, 2022).

Hypothesis 4 was regarding the textual analysis, we expected to see a more frequent use of the words "covid" during the pandemic period (P2 and P3) and a more frequent use of words "vaccination" during the period of mass vaccination (P4). Periods P2 and P3 registered a very high interaction especially for posts mentioning "covid". For period P4, we expected higher interactions for posts mentioning "vaccination", but it was not as high as for posts mentioning "covid". As we could see in Figure 14, there was not an indication of association between higher engagement and higher vaccination rate. The pre-pandemic period P1 presents a low interaction level for published posts. This might be due to the fact that the agencies didn't build trust before the pandemic, as (van den Broucke, 2021) says "health promotion should not wait until a crisis happens but prepare itself to respond swiftly."

The pandemic has made the public look for information through Facebook. However, the differences in performance from these pages may be linked to different infodemics strategies. We found 3 different types/groups/profiles of agencies.

- The first group was the one in which the global total interactions were predominant. For instance, agencies with high average total interactions (comments, shares, likes) as **Ministério da Saúde**. This might be due to the fact that Brazil is the 4th in the world of Facebook usage (Figure 13), but also that the spread of fake news using Facebook in Brazil is exorbitant (Galhardi et al., 2020).

- The second group was the one in which relative engagement was predominant. Agencies with high relative engagement performance as **Sundhedsstyrelsen** opposing agencies with low relative engagement. This may be due to the fact that the Danish community trusts their government and politicians. As (Olagnier & Mogensen, 2020) says "Denmark is a country where trust regulates everything. It is striking that Danish citizens do not see a host of conspiracy theories or widespread panic surrounding the handling of the coronavirus crisis."

- And a third group integrating agencies with high global interaction in period 3 as Ministério da Saúde and high global interactions in period 4 as Ministère des solidarités et de la santé, Ministero della Salute and DGS. This may be an indication that these agencies were the most active before and during the pandemic period.

5.2. Limitations and Future Research

This study presents a few limitations, that may represent also future paths and opportunities for research. The main limitation regards the selection of the health organizations to be studied since there are many around the globe, sometimes more than one within each country, and they have different engagement performances with the public. Further research is needed to explore their strategies in health promotion Facebook communication approach. Here we may analyze the engagement performance of health agencies but cannot assess the success of their strategies, since we lack the information to do so.

Another limitation is that the focus of the study was only on Facebook's performance. There are different social media platforms, and the use of other platforms evolves differently throughout time, countries, population characteristics, etc (Figure 12). Further research is needed to compare engagement between different platforms and link it to different population traits.

Additionally, a data source limitation is to be expected in the future. After this study, there have been reports of divestment by Meta (Facebook) in the CrowdTangle platform, which may hinder a more comprehensive analysis in the future. It is important to highlight the importance of ensuring that researchers have access to this information continuously over time, in an infodemics context.

Moreover, only publicly available Facebook KPIs were analyzed here, focusing on organic engagement. The page owners can, and should, have more information that could generate further insights. This additional information is especially important to determine the kind of content that generates the most engagement, positively.

To finish, sentiment analysis was not performed to distinguish between positive and negative engagement. Further research is now underway to perform text analysis, especially in a context where different languages are being used to disseminate a similar message.

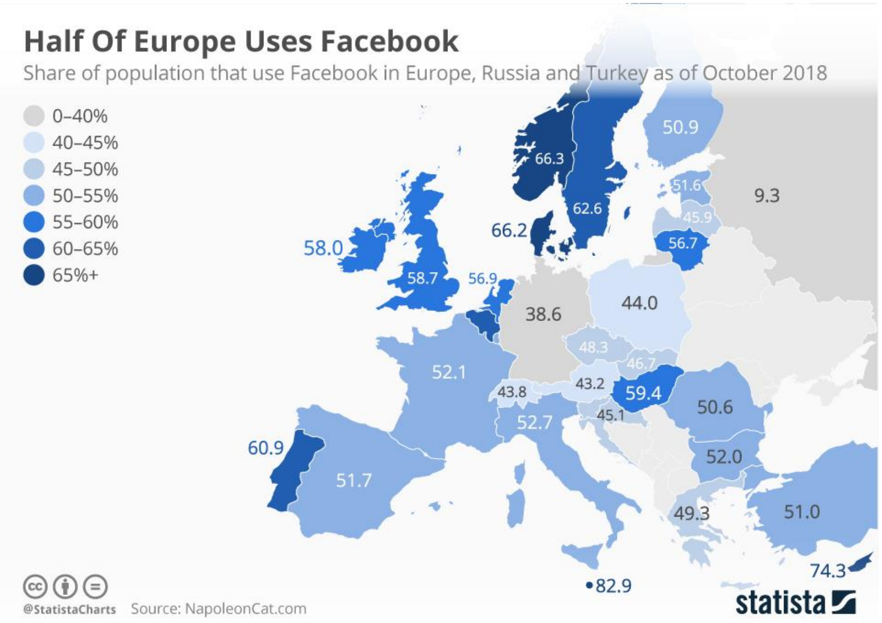


Figure 12 - Percentage of total Facebook users per country (statistica chart https://www.statista.com/chart/16256/facebook-users-in-europe/)

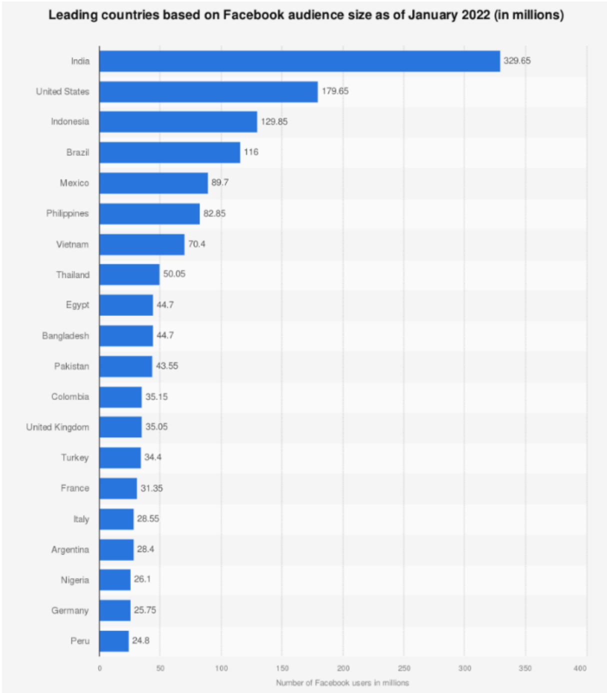
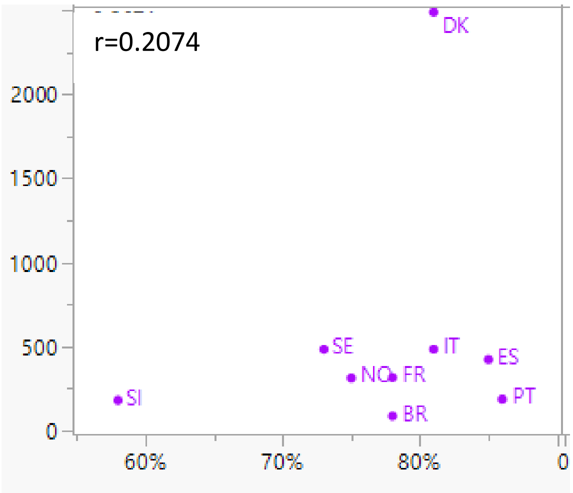


Figure 13 – Facebook usage around the world in millions

(https://ec.europa.eu/eurostat/databrowser/view/tin00127/default/map?lang=en)

Figure 14 - Scatterplot of the Mean of Relative Engagement during P4 (axis y) and vaccination rate by the end of P4 (with the correlation coefficient r=0.2074) (axis x)



6. CONCLUSIONS

In this study, we characterized the individual engagement performance of social media posts published on Facebook pages of selected national health organizations before and during the Covid19 pandemic. We also presented a typology of Facebook's performance of national health organizations during the pandemic period.

We found that the COVID -19 pandemic and social media platforms such as Facebook enabled national health agencies to more directly reach populations interested in health information. In addition, our results show that the pandemic represented a shift in how people interact and engage with health agencies on social media, particularly on Facebook, as the public did indeed turn to social media for information. However, some health organizations may not have taken enough advantage of social media to engage with their current and potential users.

In summary, the infodemic management should not end after the crisis has been averted or has disappeared from the public discussion but should be an ongoing investment. This may represent one of the best ways to make a more effective and competent health promotion.

7. REFERENCES

Alonso-Cañadas, J., Galán-Valdivieso, F., Saraite-Sariene, L., & Caba-Pérez, C. (2020). Committed to Health: Key Factors to Improve Users' Online Engagement through Facebook. *International Journal of Environmental Research and Public Health*, *17*(6), 1814. https://doi.org/10.3390/ijerph17061814

Anderson, L., Bell, H. G., Gilbert, M., Davidson, J. E., Winter, C., Barratt, M. J., Win, B., Painter, J. L., Menone, C., Sayegh, J., & Dasgupta, N. (2017). Using Social Listening Data to Monitor Misuse and Nonmedical Use of Bupropion: A Content Analysis. *JMIR Public Health and Surveillance*, *3*(1), e6. https://doi.org/10.2196/publichealth.6174

Bel Trew. (2020). *Coronavirus: Hundreds dead in Iran from drinking methanol amid fake reports it cures disease*. https://www.independent.co.uk/news/world/middle-east/iran-coronavirus-methanol-drink-cure-deaths-fake-a9429956.html

Bernardino, M., & Bacelar Nicolau, L. (2020). The importance of reliable social media information during the COVID-19 pandemic. *European Journal of Public Health*, *30*(Supplement_5). https://doi.org/10.1093/eurpub/ckaa165.067

Calleja, N., AbdAllah, A., Abad, N., Ahmed, N., Albarracin, D., Altieri, E., Anoko, J. N., Arcos, R., Azlan, A. A., Bayer, J., Bechmann, A., Bezbaruah, S., Briand, S. C., Brooks, I., Bucci, L. M., Burzo, S., Czerniak, C., de Domenico, M., Dunn, A. G., … Purnat, T. D. (2021). A Public Health Research Agenda for Managing Infodemics: Methods and Results of the First WHO Infodemiology Conference. *JMIR Infodemiology*, *1*(1), e30979. https://doi.org/10.2196/30979

Centre for Disease Prevention, E. (n.d.). *Risk of COVID-19 transmission related to the end-of-year festive season*.

Ding, K., Shu, K., Li, Y., Bhattacharjee, A., & Liu, H. (2020). *Challenges in Combating COVID-19 Infodemic -- Data, Tools, and Ethics*. http://arxiv.org/abs/2005.13691

Doj, C. (2020). *Anti-Asian Hate Crime Events During the COVID-19 Pandemic - Report - CJSC Publications - California Department of Justice*. https://www.pbs.org/newshour/nation/we-have-been-through-this-before-why-

DON NUTBEAM. (2006). Health literacy as a public health goal: a challenge forcontemporary health education and communication strategies into the 21st century. *Department of Public Health and Community Medicine,*.

Donovan, J. 2020. (2020). *Here's how social media can combat the coronavirus 'infodemic'.* https://www.technologyreview.com/2020/03/17/905279/facebook-twitter-social-media-infodemic-misinformation/

Ecdc. (2022). *Overview of the implementation of COVID-19 vaccination strategies and vaccine deployment plans in the EU/EEA April 2022*.

ELLA HOLLOWOOD, & ALEXI MOSTROUS. (2020). *Fake news in the time of C-19*. https://www.tortoisemedia.com/2020/03/23/the-infodemic-fake-news-coronavirus/

Eysenbach, G. (2002). Infodemiology: the epidemiology of (mis)information. *The American Journal of Medicine*, *113*(9), 763–765. https://doi.org/10.1016/S0002-9343(02)01473-0

Eysenbach, G. (2003). SARS and population health technology. In *Journal of Medical Internet Research* (Vol. 5, Issue 2, pp. 1–8). JMIR Publications Inc. https://doi.org/10.2196/jmir.5.2.e14

Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. In *Journal of Medical Internet Research* (Vol. 11, Issue 1). JMIR Publications Inc. https://doi.org/10.2196/jmir.1157

Eysenbach, G. (2020). How to fight an infodemic: The four pillars of infodemic management. In *Journal of Medical Internet Research* (Vol. 22, Issue 6). JMIR Publications Inc. https://doi.org/10.2196/21820

Galhardi, C. P., Freire, N. P., Minayo, M. C. de S., & Fagundes, M. C. M. (2020). Fato ou Fake? Uma análise da desinformação frente à pandemia da Covid-19 no Brasil. *Ciência & Saúde Coletiva*, *25*(suppl 2), 4201–4210. https://doi.org/10.1590/1413-812320202510.2.28922020

Geldsetzer, P. (2020). Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19. *Journal of Medical Internet Research*, *22*(4), e18790. https://doi.org/10.2196/18790

Glik, D. C. (2007). Risk communication for public health emergencies. *Annual Review of Public Health*, *28*, 33–54. https://doi.org/10.1146/annurev.publhealth.28.021406.144123

Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Murshid Hasan, S. M., Kabir, A., Yeasmin, D., Islam, M. A., Chowdhury, K. I. A., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19-Related infodemic and its impact on public health: A global social media analysis. *American Journal of Tropical Medicine and Hygiene*, *103*(4), 1621–1629. https://doi.org/10.4269/ajtmh.20-0812

Jansen, N., Hinz, O., Deusser, C., & Strufe, T. (2021). Is the Buzz on? – A Buzz Detection System for Viral Posts in Social Media. *Journal of Interactive Marketing*, *56*, 1–17. https://doi.org/10.1016/j.intmar.2021.03.003

Kleis Nielsen, R., Fletcher, R., Newman, N., Scott Brennen, J., & Howard, P. N. (2020). *Navigating the "Infodemic": How People in Six Countries Access and Rate News and Information about Coronavirus*.

Liu, C., Wang, D., Liu, C., Jiang, J., Wang, X., Chen, H., Ju, X., & Zhang, X. (2020). What is the meaning of health literacy? A systematic review and qualitative synthesis. In *Family Medicine and Community Health* (Vol. 8, Issue 2). BMJ Publishing Group. https://doi.org/10.1136/fmch-2020-000351

Marais, F., Minkler, M., Gibson, N., Mwau, B., Mehtar, S., Ogunsola, F., Banya, S. S., & Corburn, J. (2016). A community-engaged infection prevention and control approach to Ebola. *Health Promotion International*, *31*(2), 440–449. https://doi.org/10.1093/heapro/dav003

Moon, H., & Lee, G. H. (2020). Evaluation of Korean-language COVID-19-related medical information on YouTube: Cross-sectional infodemiology study. *Journal of Medical Internet Research*, *22*(8). https://doi.org/10.2196/20775

Naeem, S. bin, Bhatti, R., & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. In *Health Information and Libraries Journal* (Vol. 38, Issue 2, pp. 143–149). Blackwell Publishing Ltd. https://doi.org/10.1111/hir.12320

Nsoesie, E. O., & Oladeji, O. (2020). Identifying patterns to prevent the spread of misinformation during epidemics. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-014

of Health Promotion Education, & Communication. (1998). *Health promotion glossary* (p. WHO/HPR/HEP/98.1). World Health Organization.

Olagnier, D., & Mogensen, T. H. (2020). The Covid-19 pandemic in Denmark: Big lessons from a small country. *Cytokine and Growth Factor Reviews*, *53*, 10–12. https://doi.org/10.1016/j.cytogfr.2020.05.005

Oyeyemi, S. O., Gabarron, E., & Wynn, R. (2014). Ebola, Twitter, and misinformation: A dangerous combination? In *BMJ (Online)* (Vol. 349). BMJ Publishing Group. https://doi.org/10.1136/bmj.g6178

Paakkari, L., & Okan, O. (2020). COVID-19: health literacy is an underestimated problem. In *The Lancet Public Health* (Vol. 5, Issue 5, pp. e249–e250). Elsevier Ltd. https://doi.org/10.1016/S2468-2667(20)30086-4

Purnat, T. D., Vacca, P., Czerniak, C., Ball, S., Burzo, S., Zecchin, T., Wright, A., Bezbaruah, S., Tanggol, F., Dubé, È., Labbé, F., Dionne, M., Lamichhane, J., Mahajan, A., Briand, S., & Nguyen, T. (2021). Infodemic Signal Detection During the COVID-19 Pandemic: Development of a Methodology for Identifying Potential Information Voids in Online Conversations. *JMIR Infodemiology*, *1*(1), e30971. https://doi.org/10.2196/30971

Purnat Tina. (2020, December 2). *Building systems for respond to infodemics and build resilience to misinformation*. https://www.linkedin.com/pulse/building-systems-respond-infodemics-build-resilience-tina-d-purnat/

Risk communication and community engagement readiness and response to coronavirus disease(COVID-19). (n.d.). Retrieved January 23, 2022, from https://www.who.int/publications/i/item/risk-communication-and-community-engagement-readiness-and-initial-response-for-novel-coronaviruses

Rubinelli, S., Purnat, T. D., Wilhelm, E., Traicoff, D., Namageyo-Funa, A., Thomson, A., Wardle, C., Lamichhane, J., Briand, S., & Nguyen, T. (2022). WHO competency framework for health authorities and institutions to manage infodemics: its development and features. *Human Resources for Health*, *20*(1), 35. https://doi.org/10.1186/s12960-022-00733-0

Simonds, S. K. (1974). Health Education as Social Policy. *Health Education Monographs*, *2*(1_suppl), 1–10. https://doi.org/10.1177/10901981740020S102

Situation Report-51 SITUATION IN NUMBERS total and new cases in last 24 hours. (n.d.).

Smith, J. A., & Judd, J. (2020). COVID-19: Vulnerability and the power of privilege in a pandemic. *Health Promotion Journal of Australia*, *31*(2), 158–160. https://doi.org/10.1002/hpja.333

Tangcharoensathien, V., Calleja, N., Nguyen, T., Purnat, T., D'Agostino, M., Garcia-Saiso, S., Landry, M., Rashidian, A., Hamilton, C., AbdAllah, A., Ghiga, I., Hill, A., Hougendobler, D., van Andel, J., Nunn, M., Brooks, I., Sacco, P. L., de Domenico, M., Mai, P., … Briand, S. (2020). Framework for managing the COVID-19 infodemic: Methods and results of an online, crowdsourced who technical consultation. *Journal of Medical Internet Research*, *22*(6). https://doi.org/10.2196/19659

Tasnim, S., Hossain, M., & Mazumder, H. (2020). Impact of rumors and misinformation on COVID-19 in Social Media. In *Journal of Preventive Medicine and Public Health* (Vol. 53, Issue 3, pp. 171–174). Korean Society for Preventive Medicine. https://doi.org/10.3961/JPMPH.20.094

Timothy Caulfield. (2020). *Pseudoscience and COVID-19 — we've had enough already*. https://www.nature.com/articles/d41586-020-01266-z

Tran, T., & Lee, K. (2016). Understanding citizen reactions and Ebola-related information propagation on social media. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 106–111. https://doi.org/10.1109/ASONAM.2016.7752221

van den Broucke, S. (2021a). Why health promotion matters to the COVID-19 pandemic, and vice versa. In *Health Promotion International* (Vol. 35, Issue 2, pp. 181–186). Oxford University Press. https://doi.org/10.1093/HEAPRO/DAAA042

van den Broucke, S. (2021b). Why health promotion matters to the COVID-19 pandemic, and vice versa. In *Health Promotion International* (Vol. 35, Issue 2, pp. 181–186). Oxford University Press. https://doi.org/10.1093/HEAPRO/DAAA042

WHO. (2020). *Situation Report-51 SITUATION IN NUMBERS total and new cases in last 24 hours*.

WHO competency framework Building a response workforce to manage infodemics. (2021).

Zarocostas, J. (2020). How to fight an infodemic. *Lancet (London, England)*, *395*(10225), 676. https://doi.org/10.1016/S0140-6736(20)30461-X