Università
Ca'Foscari
Venezia

Master's degree
in Management

Final Thesis

The relationship between the application of Data Science and the optimization of digital marketing campaigns

An empirical investigation in Twitter

Supervisor
Prof. Carlo Romano Marcello Alessandro Santagiustina

Graduand
Marziye Goodarzi
Matriculation number
888193

Academic Year
2022 / 2023

Acknowledgments

First and foremost, I want to give a big thank you to my thesis supervisor, Professor Carlo Romano Marcello Alessandro Santagiustina, for their invaluable guidance and support.

I also want to thank all the faculty, staff, and fellow students at the University of Venice (Ca' Foscari). You all made my time here awesome and I'm grateful for your support.

And of course, a special thanks to my family and friends for always being there for me. Your love and encouragement meant the world to me.

Index

Introduction

With the growth of the internet and social media, digital marketing has become a critical aspect of any business's marketing strategy. Companies are investing more resources in digital marketing to reach their target audience and interact with customers. Social media platforms like Twitter have emerged as important channels for digital marketing, offering companies the ability to interact directly with users and potential customers. However, with so much information available, it can take time to determine how to create effective campaigns that resonate with users. In recent years, the application of data science has become increasingly prevalent in digital marketing. One area of particular interest is the optimization of digital marketing campaigns using data-driven insights.

Data science is an interdisciplinary field that involves the extraction, processing, analysis, and interpretation of large and complex datasets to identify patterns, trends, and insights. In the context of digital marketing, data science provides companies with the ability to analyze large datasets from social media platforms to understand user behavior, preferences, and opinions. This information can be used to optimize digital marketing campaigns and improve the effectiveness of a company's communication strategy.

Using a hybrid framework of text mining and structural topic modeling (STM), this study aims to investigate how users engage with the advertisement campaigns on Twitter and how the product identity communicated in the campaign aligns with what users communicate about the product. In addition, the study seeks to provide insights into measuring the success of online advertising of a product through social media and identifying the terms and topics that users associate with the advertisement campaigns. By achieving these objectives, this study will contribute to the expanding literature on data science in social media data analysis and provide practical implications for companies looking to optimize their digital marketing campaigns.

As a case study, Samsung Mobile has an active presence on social media platforms, including Twitter, and uses these channels to promote its products and communicate with its customers. By analyzing the advertisement campaign contents of tweets referring to Samsung Mobile's #GalaxyS21 product, this study aims to provide insights into the relationship between the application of data science and the optimization of digital marketing campaigns.

Research Questions

The research questions for this study are as follows:

- To what extent is the product identity communicated in advertisement campaigns proposed by Samsung Mobile on their social media profiles aligned with what users communicate about the product?
- Was Samsung Mobile's online communication for the Galaxy S21 product successful?
- How can we measure the success of online advertisements of a product through social media?
- How and in which terms are users aligned to the advertisement campaigns of Galaxy S21?

These research questions are designed to investigate the effectiveness of Samsung Mobile's online communication strategy for the Galaxy S21 product and provide insights into how users respond to the product identity communicated in the advertisement campaigns proposed by Samsung Mobile.

Objectives of the Study

The objectives of this study are:

- To investigate the alignment between the product identity communicated in the Samsung Mobile #GalaxyS21 advertisement campaign and what users communicate about the product.
- To assess the effectiveness of Samsung Mobile's online communication for the Galaxy S21 product.
- To provide insights into how to measure the success of online advertisement of a product through social media.
- To identify the terms and topics that users associate with the advertisement campaigns of Galaxy S21.

To achieve these objectives, a hybrid framework of text mining and structural topic modeling techniques will be used to analyze the contents of tweets referring to Samsung Mobile's #GalaxyS21 advertisement campaign.

Significance of the Study

The significance of this study lies in its contribution to analyzing the alignment between Samsung Mobile's online communication and its users through the use of a hybrid framework of text mining and structural topic modeling (STM). By examining the advertisement campaign's contents and the topics that users associate with the campaign, this study seeks to provide insights into how well Samsung Mobile's online communication strategy aligns with its users' perception of the product. The findings of this study can help Samsung Mobile and other companies to optimize their digital marketing campaigns by providing insights into how to communicate with their target audience effectively.

This thesis is structured as follows: In Chapter 1 literature review is given in three different sub-sections. In Chapter 3, the research design, data collection methods, and techniques are described. The results and analysis are presented in Chapter 4. In Chapter 5, we discuss our findings around research questions and objectives. Finally, concluding remarks are presented in Chapter 6.

Chapter 1

Literature Review

In our thesis research, we have investigated several literature streams that are vital to our research objectives. Our initial literature stream focuses on the communication of product identity for high-tech products on social media platforms. We examined studies that explore the influence of social media on consumers' purchasing decisions, brand perceptions, and engagement with advertisements. This literature is crucial as it provides a foundation for our research, helping us understand the broader context in which Samsung Mobile's online communication efforts operate. Another important stream of literature revolves around the effectiveness of online advertising campaigns. We examined various metrics used to assess the success of these campaigns, such as reach, engagement, conversions, and return on investment. Additionally, we explored research on targeting strategies, creative content, and campaign optimization techniques. Lastly, we delved into literature related to the analysis of online communication patterns. We explored different methodologies, including sentiment analysis, natural language processing, and network analysis, to uncover insights from user-generated content. By studying this literature, we aim to decipher the sentiments, opinions, and topics prevalent in Samsung Mobile's online communication and identify potential gaps or inconsistencies.

1.1. Online Communication in social media Related to High-Tech Products and the Communication of Their Identity

In recent years, social media has become a significant channel for companies to communicate with their customers. Research has shown that companies can effectively communicate their product identity by leveraging the features of social media platforms. For example, Twitter allows companies to communicate with their customers in real-time and enables users to engage with the brand by using hashtags or mentions (Culnan et al., 2010).

Social media platforms provide a unique opportunity for firms to create brand awareness, build customer loyalty, and reach out to potential customers (Kaplan and Haenlein, 2010). One of the critical aspects of online communication is how firms communicate the identity of their products to customers. The identity of a product refers to its unique features and benefits, which distinguishes it from other similar products in the market (Ghazizadeh et al., 2015). In the context of high-tech products, such as smartphones, the communication of product identity is of utmost importance, as customers tend to make purchase decisions based on the perceived value and uniqueness of the product (Bian & Wang, 2017). In other words, social media can be particularly effective for high-tech products as they are often the subject of intense interest and discussion among tech-savvy consumers.

Several studies have investigated the communication of product identity on social media platforms for high-tech products. Chen et al. (2018) examined the relationship between brand personality, social media engagement, and purchase intentions for high-tech products on Facebook. The results showed that brand personality and social media engagement positively influenced purchase intentions. Similarly, Jiang et al. (2019) investigated how brand identity on social media influenced customers' purchase intentions for smartphones. They found that brand identity positively influenced customers' purchase intentions, and emotional attachment to the brand played a mediating role.

In particular, the alignment between a firm's messaging and the perceptions of its customers is critical in building trust and credibility. One study that explored this issue in the context of social media was conducted by Gupta and Kim (2021). They analyzed tweets related to Apple's iPhone 11 product launch and found that the alignment between the product's identity communicated by Apple and the identity perceived by customers was positively associated with their purchase intentions. Similarly, another study by Zhang and Mao (2021) found that the online communication of product identity was positively related to customer engagement and loyalty.

Several studies have examined the effectiveness of online product advertisement campaigns. For example, Godey et al. (2016) investigated the impact of social media advertising on the customer's decision-making process and found that social media advertising significantly influences the customer's attitude towards the product and purchase intention. Similarly, Li and Liang (2019) analyzed the effectiveness of mobile social media advertising and found that the use of visually appealing and personalized content significantly increases the engagement rate and positively impacts the customer's attitude towards the product.

1.2. Measuring the Success of Online Product Advertisement Campaigns

A study by Kalyanaraman and Sundar (2006) found that measuring the effectiveness of online advertising is challenging due to the difficulty of tracking online user behavior. However, the study also noted that measuring user engagement and feedback can provide valuable insights into the success of online product advertisement campaigns. A study by Lu, Hsiao, and Chiu (2017) found that social media analytics can be used to effectively measure the success of online product advertisement campaigns by analyzing user engagement, sentiment, and reach. It allows them to understand the impact of their messaging and optimize their strategies accordingly.

In recent years, a wide range of tools and metrics have been developed to help companies track and measure the effectiveness of their online advertising campaigns.

One of the most commonly used metrics for measuring the success of online advertising campaigns is click-through rate (CTR). CTR measures the percentage of users who click on an ad after being exposed to it and is calculated by dividing the number of clicks by the number of impressions. While CTR is a widely used metric, it has been criticized for its narrow focus on the number of clicks, which may not always correlate with actual conversions or sales (Chaffey & Ellis-Chadwick, 2019). The second widely used metric to measure the success of online advertising campaigns is engagement rate, which measures the level of interaction between customers and the advertisement content (Godey et al., 2016). The engagement rate includes various types of interactions, such as likes, shares, comments, and clicks.

Another important metric for measuring the success of online advertising campaigns is conversion rate. Conversion rate measures the percentage of users who take a desired action after clicking on an ad, such as making a purchase or filling out a form. This metric provides a more accurate measure of the effectiveness of an advertising campaign than CTR, as it focuses on actual customer behavior rather than just clicks. Companies can track conversion rates using tools such as Google Analytics or similar web analytics platforms (Chaffey & Ellis-Chadwick, 2019). In addition, other metrics are commonly used to measure the success of online advertising campaigns, including cost per acquisition (CPA), return on investment (ROI). CPA measures the cost of acquiring a new customer or lead, while ROI measures the financial return on an advertising campaign.(Chaffey & Ellis-Chadwick, 2019).

However, these metrics may not provide a complete picture of the impact of the campaign on customer attitudes and behaviors. Advancements in data science have provided a variety of tools and techniques to measure the effectiveness of online product advertisement campaigns.

Recent research has suggested that social media metrics such as sentiment analysis can provide a more comprehensive view of campaign success (Kietzmann et al., 2011; Vashisht et al., 2019). For example, a study by Li and Liang (2021) analyzed tweets related to Nike's "Just Do It" campaign and found that sentiment analysis provided a more accurate view of the impact of the campaign on customer attitudes than click-through rates.

Sentiment analysis is a popular tool for measuring the success of online product advertisement campaigns. It is a technique used to extract and analyze the sentiment of online user reviews, comments, and feedback regarding a particular product or service. The sentiment analysis tool assigns a positive or negative score to a given text or review. Several studies have used sentiment analysis to measure the success of online product advertisement campaigns. For instance, the study by Xiong et al. (2017) analyzed the online customer reviews of the iPhone 6S on Amazon and found that the positive sentiment of the reviews was strongly related to sales. (Gutnik, 2021) applied Data Mining and Machine Learning Methods to enhance the effectiveness of digital marketing strategies in the IT sector. The author focused on machine learning algorithms (clustering, regression, classification, neural networks, decision trees). Using sentiment analysis and topic extraction, Ektoros et al. (2019) examined customer sentiment in tweets on the release of two smartphone models. Data from Twitter was obtained. The research question addressed in this study was if the company under investigation's marketing positioning approach was successful following the release of two new devices, the Huawei P20 and Huawei P20 Pro. They discovered evidence that Huawei's marketing strategy had a favorable influence on its target users. The firm's product positioning strategy was to distinguish on pricing and technological attributes. According to preliminary data, buyers liked the new triple camera capability of the P20 and mentioned it in good tweets. Using a Naïve Bayes classifier with a corpus-based approach (Cassone et la., 2020) applied sentiment analysis to analyze some branded hashtag campaigns in Twitter. The outcome of their study was an assessment of the campaigns by evaluating the customer opinions and gathering the most used words from the customers and their geolocations.

Text mining is another technique used to measure the success of online product advertisement campaigns. Text mining involves the extraction of relevant information and patterns from large amounts of unstructured data. This technique is applied to analyze online user-generated content such as product reviews, comments, and feedback. In a study by Godes and Mayzlin (2004), text mining was used to analyze user-generated content on Epinions.com to measure the effect of word-of-mouth on product sales. The study found that the volume of user-generated content had a positive impact on product sales.

Topic modeling is a technique used to identify the underlying themes or topics in a large corpus of text. It is used to extract relevant information from unstructured text data, which can be used to measure the success of online product advertisement campaigns. Several studies have used topic modeling to analyze user-generated content on social media platforms to measure the success of product advertisement campaigns. For example, the study by Cao et al. (2017) used topic modeling to analyze user-generated content on Twitter to measure the effectiveness of the advertising campaign for Coca-Cola. The study found that the advertising campaign generated a positive impact on user engagement and online sentiment.

There has been an increase in interest among marketing researchers and practitioners in using topic models in a variety of marketing application areas (Reisenbichler and Reutterer, 2018). Using topic modeling and Decision Tree (DT) models Gregoriades et al. (2021) proposed a machine learning approach to optimize marketing campaigns, that is, the communication of the right content to the right consumers. They focused on the tourist industry. They suggested this approach to build messages that better target the goals and requirements of tourists from various cultural and economic backgrounds. They applied topic modeling to identify the major concepts discussed in eWOM and decision trees for pattern recognition using rules extraction.

1.3. How to analyze online communication through STM

In the past few years, there has been an increase in interest among marketing researchers and practitioners in using topic models in a variety of marketing application areas (Reisenbichler and Reutterer, 2018). Topic modeling is a specific text mining technique used to uncover hidden themes or topics within a collection of documents. It is an unsupervised machine learning approach that automatically identifies patterns of word co-occurrences in the text to generate a set of topics. These topics represent the main themes present in the documents and help in understanding the underlying structure of the text corpus (Blei et el., 2003).

Using topic modelling and Decision Tree (DT) models Gregoriades et al. (2021) proposed a machine learning approach to optimize marketing campaigns, that is, the communication of the right content to the right consumers. They focused on tourist industry. They suggested this approach to build messages that better target the goals and requirements of tourists from various cultural and economic backgrounds. They applied topic modelling for identifying the major concepts discussed in eWOM, and decision trees for patterns recognition using rules extraction.

Structural Topic Modeling (STM) is an extension of traditional topic modeling techniques that incorporates metadata and document structure to uncover latent thematic patterns within a text corpus. While traditional topic models such as Latent Dirichlet Allocation (LDA) focus solely on identifying topics based on word co-occurrences, STM takes into account additional information, such as document metadata (e.g., author, publication date) and document structure (e.g., sections, paragraphs).

In STM, the topics are not only defined by the distribution of words but also by the distribution of metadata variables. This allows for a more nuanced analysis, as it enables researchers to examine how different metadata factors influence the prevalence of certain topics within the corpus. For example, in a corpus of news articles, the metadata variables could include the newspaper source, publication date, or the section of the newspaper where the article appears. By incorporating these metadata variables, STM can uncover topics that are specific to certain sources or time periods.

STM employs a probabilistic model that combines a topic model with regression models. It simultaneously estimates the topic proportions within documents and the relationships between topics and metadata variables. This allows researchers to explore how the occurrence of topics varies across different metadata categories and to identify the most influential metadata variables on topic prevalence (Roberts et al., 2016).

One significant advantage of employing structural topic modeling is the ability to integrate metadata as variables influencing topic proportions and content. This allows for the recording of differences in how (topic contents) and how much (topic proportions) groups speak about the various predicted outcomes. These distinctions are not captured by traditional topic modeling approaches such as Latent Dirichlet Allocation (LDA) or the Correlated Topic Model (CTM) (Santagiustina and Warglien, 2022).

Barry et al. (2016) used cutting-edge, impartial content analysis to investigate the Twitter advertising strategies of top alcoholic beverage firms. They examined the whole history of 13 alcoholic beverage brands' tweets using Latent Dirichlet allocation (LDA). For each brand, unique, diverse concepts are developed. Each brand had its own marketing strategy that reflected the brand's personality. He et al. (2020) used the structural topic model (STM) to extract hidden topics from web-based drug reviews and looked at customers' worries about online drug purchases. They identified 12 topics. They concluded five negative topics from the evaluations which signify consumer dissatisfaction. Ibrahim and Wang (2019) concentrated on analyzing brand-related tweets connected to five top UK online retailers during the crucial holiday shopping season. Utilizing a variety of data analytics techniques, including as time series analysis, sentiment analysis, and topic modeling, they investigated patterns in customer tweets in order to analyze the trends of tweet volume and sentiment and to comprehend the factors that underlie changes in sentiment. To further understand what precisely causes these shifts in sentiment, they utilized a topic modeling method to analyze the tweets that were sent before and after these pivotal times. They concluded by offering a way for businesses to evaluate their performance and prioritize the areas that need more of their focus.

While our work shares similarities with the previously mentioned literature in terms of exploring the impact of social media on consumer behavior, there are several key differences in terms of research design.

Firstly, unlike the existing literature that focuses on consumer behavior and online advertising campaigns, our study specifically hones in on the alignment between Samsung Mobile's online communication and user-generated content through STM. This research design difference allows us to delve deeper into the specific context of Samsung Mobile's communication efforts and analyze the extent to which the product identity communicated in their advertisement campaigns aligns with users' perceptions.

Secondly, our study incorporates STM technique. This methodology enable us to extract insights from the vast amount of user-generated content on social media platforms, decipher sentiments, opinions, and topics related to Samsung Mobile's products, and identify potential gaps or inconsistencies in their communication strategy. By employing these research design differences, we aim to answer unique questions that have not been thoroughly explored before, such as the alignment between Samsung Mobile's communication and user perceptions and the effectiveness of their online communication in shaping brand image.

The importance of these research design differences lies in their ability to provide a more comprehensive understanding of the specific communication dynamics between Samsung Mobile and its users. This allows us to address specific gaps in knowledge, identify areas for improvement, and provide valuable recommendations for enhancing Samsung Mobile's online communication strategies.

Chapter 2

Data and Methods

2.1. Case Study

The case study focuses on Samsung Mobile's communication and feedback effects on Twitter regarding the Galaxy S21.

The Galaxy S21, a flagship product unpacked on January 14, 2021, and released on 29 January 2021 holds significance as the focal point of the analysis. As a widely discussed device, it is likely to have generated substantial user engagement on Twitter.

By delving into Samsung's communication strategies on Twitter, we can gain comprehensive insights into how the company engages with its clients. Analyzing the content of Samsung's tweets and audience tweets enable an examination of the tone and overall communication approach employed by the company. Understanding these aspects helps to decipher how Samsung manages its brand image, promotes its products, and addresses customer queries or concerns.

Furthermore, the study offers an opportunity to explore the feedback effects that emerge from customer interactions on Twitter. Twitter provides a public platform where users openly express their opinions, experiences, and suggestions regarding products and brands. Analyzing Samsung's communication strategies on Twitter allows us to assess the ways in which customer input shapes the company's marketing strategies. This investigation contributes to a deeper understanding of how customer opinions align with the company's communication strategies.

Twitter's role as a communication channel adds another layer of interest to this case study. With its real-time nature and features such as hashtags, retweets, and mentions, Twitter facilitates direct and public interactions between brands and customers. This unique platform enables an exploration of the dynamics of customer engagement, sentiment, and brand advocacy. Analyzing Twitter data related to Samsung Mobile and the Galaxy S21 provides valuable insights into the effectiveness of the company's communication strategies and its ability to connect with customers in a public and open environment.

Moreover, the selection of Samsung as the subject of this case study is significant due to the brand's global recognition and dominant position in the mobile industry. Samsung's extensive experience and expertise in marketing mobile devices make it an excellent source for identifying successful communication practices and establishing benchmarks for evaluating communication efforts within the IT sector.

2.2. Twitter data collection

The data collection for this study focuses on gathering tweets that mention the hashtag #GalaxyS21 from January 2021 to December 2022, spanning a period of two years. This duration allows for a comprehensive analysis of the marketing efforts associated with the Samsung #GalaxyS21 over a significant timeframe. Twitter's full-Archive Search API V2 is utilized to collect the data. The R Studio software is used for analyzing data.

To distinguish between the communication by the company and that of the users, the collected data is categorized accordingly. Specifically, tweets originating from the Samsung Mobile international accounts are identified as company communication. In the next sub-section, we will explain more about this process.

2.3. Data cleaning

Following the data collection phase, the downloaded tweets undergo pre-processing steps to ensure the cleanliness and relevance of the data. This preprocessing stage involves the removal of posts that are unrelated to the subject matter, thereby filtering out any tweets that do not pertain to the Samsung #GalaxyS21. For this purpose, any irrelevant character that lacks meaning, such as emojis, spaces, URLs, etc., has been removed from the expressions. Characters such as commas (,) or any other cases that should not be used in creating bigrams have been eliminated from the data, and the final dataset has been prepared and stored. Furthermore, the tweets have been tokenized, and unintentional commas have been removed. Tokenization ensures that words used in combination are processed separately. To address this issue, after tokenization, the Bigram method has been used. This dataset is then used for further analysis.

Then, we categorized the data based on individuals' IDs. To distinguish official Samsung accounts from others, the categorization process was implemented. By applying specific criteria, such as account verification or identifying official account indicators, we were able to differentiate between tweets posted by official Samsung accounts and those from other users.

Using a regular expression we have identified multiple Samsung accounts, these accounts are both "verified" and contain the regular expression (RegEx) "samsung", with any combination of lowercase and uppercase characters. The Samsung accounts so identified are the following:

"SamsungGulf" "Samsung Mobile" "Samsung Business USA" "Samsung Mobile US" "Samsung Indonesia" "SAMSUNG DEVELOPERS" "Samsung US Newsroom" "Samsung UK" "Samsung Electronics" "Samsung Egypt" "Samsung Caribbean" "Samsung Philippines" "Samsung India" "Samsung Switzerland" "Samsung Mobile SA" "Samsung Mobile Kenya" "Samsung Mobile NG" "Samsung Mobile TZ" "SamsungNewsroomIN" "Samsung Ghana" "Samsung Pakistan" "Samsung España" "Samsung Malaysia" "Samsung Thailand" "Samsung Ireland" "Samsung SriLanka" "SamsungSA" "Samsung Networks" "Samsung Chile" "Samsung Levant" "Samsung Nepal" "Samsung Support US" "Samsung Maroc" "Samsung TV Plus" "Samsung Healthcare" "Samsung FR" "Samsung Canada" "Samsung Saudi Arabia" "Samsung Australia" "Samsung Tunisie". On the other hand, user-generated content is labeled as user communication.

In the studied data, there were 127,006 data points related to the year 2021 and 6,963 data points related to the year 2022. To summarise the characteristics of the dataset, Figure 1 and Figure 2 demonstrate the tweet frequency by date, and the average number of likes per tweet of official Samsung Mobile accounts for each month of the years 2021 and 2022, respectively.

Figure 1. The tweet frequency by date

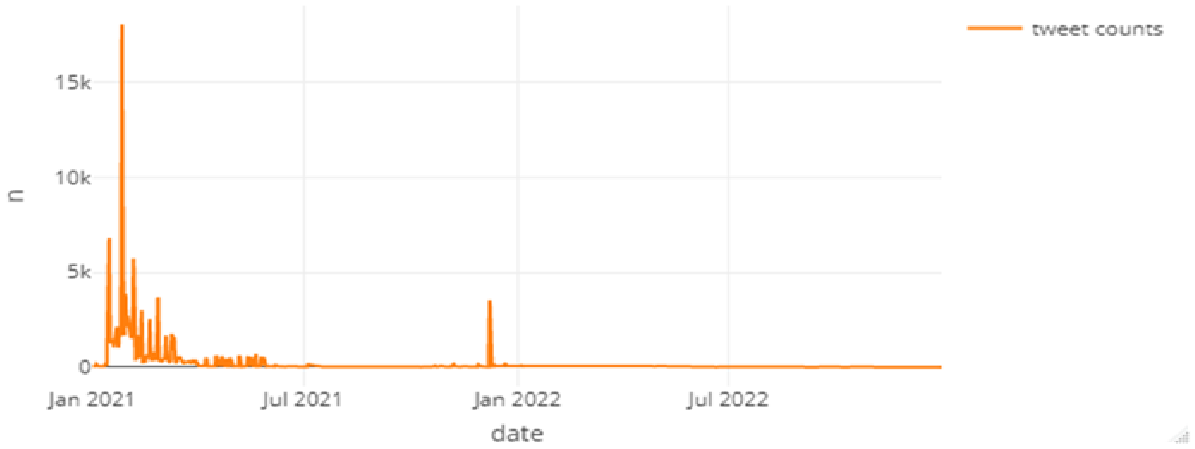


Figure 1 depicts the frequency of tweets containing the hashtag #GalaxyS21 over a period of two years. The graph showcases notable fluctuations in tweet volumes in the first 6 months of 2021, which align with the product's announcement and launch from the 14th to the 29th of January (unpacked on January 14, 2021, and released on 29 January 2021). Several dates stand out due to a significant surge in tweet activity, such as January 15th, January 26th, and February 5th. More than 18000 tweets were generated on January 26th.

Figure 2. The average number of likes per tweet

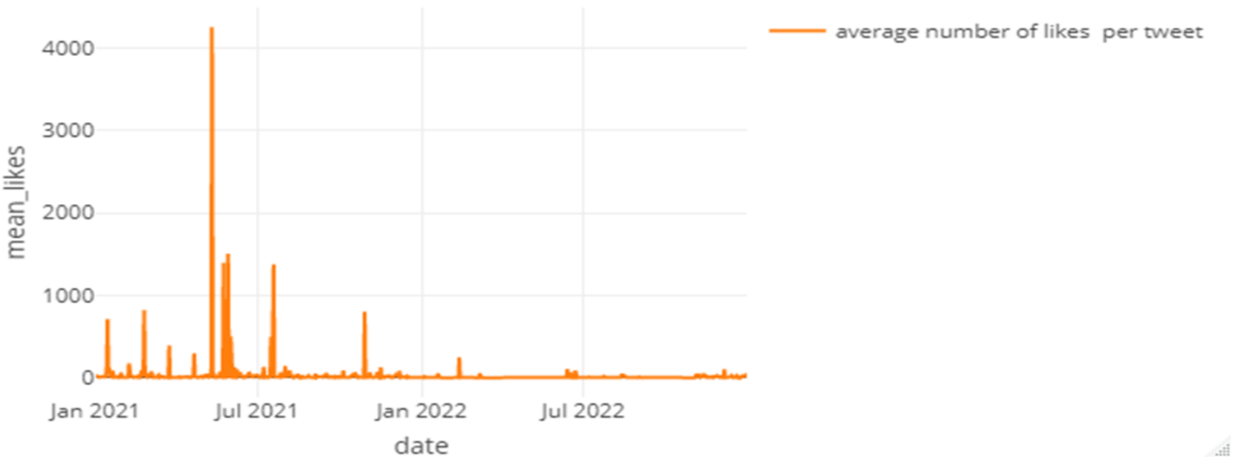


Figure 2 showcases the average number of likes per tweet featuring the hashtag #GalaxyS21 over a span of two years. Notably, on May 11th, 2021, the average number of likes per tweet soared, surpassing 4000. This spike in engagement could be attributed to an event that occurred in May 2021 when Samsung released an Android security patch for the Galaxy S21 series. This update included enhancements to the camera performance of the phones in the series as well as improvements to the Quick Share functionality. User @AdamleeTs92 reported this update through a tweet, which likely contributed to the significant number of likes observed in May 2021.

In 2021, Samsung Mobile started a partnership with BTS, a globally renowned South Korean music group, which has a notable history in relation to the Galaxy S21. The collaboration began in early 2021 when Samsung appointed BTS as its global brand ambassador. BTS actively participated in various marketing activities and promotions for the Galaxy S21.

In the next chapter, we will show to what extent these marketing activities were successful by analyzing the content of the tweets and delving into the alignment between Samsung Mobile's communication and user engagement.

A framework of text mining and structural topic modeling is used to analyze advertisement campaign contents of tweets referring to Samsung mobile (#GalaxyS21).

2.4. Method

In this section, first a brief examination of methods and tools for text mining is presented, particularly in the context of social sciences. With the advancements in computer technology, text analysis has become more accessible and beneficial in various fields. Traditionally, conveying information from individuals in a quantifiable manner was challenging. Close-ended surveys were used, but they limited the subject's discretion in answering. Extracting information from open-ended answers was nearly impossible due to the sheer volume of sentences and the need for objective classification.

One effective application of computer-assisted methods is Sentiment Analysis, which uses dictionaries to classify documents as positive, neutral, or negative based on specific words. More advanced applications can identify a broader range of human sentiments such as anger, sadness, or anxiety. This method shares similarities with the computation of the EPU (Economic Policy Uncertainty) index mentioned earlier. Another area of expansion in text mining is topic models, which will be discussed in detail in the subsequent pages.

2.4.1. Topic Models

Topic models derive their name from their underlying assumption that within a corpus of documents, there exist hidden topics. These models posit that the observable data is generated based on the joint probability of variables that represent these topics (Wesslen, 2018). In computer-based text analysis, there are two main approaches: Natural Language Processing (NLP) and statistical-based algorithms. The primary distinction between these approaches lies in their focus. NLP emphasizes interpretation by analyzing the grammatical structure of the text, while statistical algorithms often disregard word order and adopt a "bag-of-words" approach.

In the "bag-of-words" approach, the text is transformed into a document-term matrix, where each row represents a document, and each column represents a word. The entries in the matrix indicate the frequency of a particular word within a document. The Latent Dirichlet Allocation (LDA), developed by Blei et al. (2003), is a prominent example of such models. LDA is a generative probabilistic model that represents documents as random combinations of latent topics, with each topic characterized by a distribution over words.

Further advancements have been proposed by Mimno and McCallum (2008), introducing the possibility of incorporating covariates that influence topic prevalence. This means that the prevalence of topics in a document can be influenced by additional sample independent variables or metadata. Eisenstein et al. (2011) focused on the content of topics, specifically examining the word composition of each topic, and allowing for the potential influence of covariates.

In this thesis, the chosen model is the Structural Topic Model (Roberts et al., 2013), which incorporates all the aforementioned features while also introducing new functions.

2.4.2. Structural Topic Model

To gain an understanding of the model, we will provide a description of its components. We have a collection of documents, the Twitter posts, known as a corpus, where each document is indexed by $d \in \{1 . . . D\}$, representing D_d documents:

$$\{D_1, D_2 . . . D_D\}$$

Each document consists of words, and the model keeps track of their positions within the document. The position of words is indexed by $n \in \{1 . . . N_d\}$, meaning that for each word $w_{d,n}$ it is known the document to which it is in and its position. All words are assumed to belong to a vocabulary indexed by V_v, with a total size of V. Therefore, the vector $\{V_1, V_2 . . . V_V\}$ comprises unique instances of all the words used in the corpus.

The model assumes that each document is a combination of topics, where a topic represents a mixture of words. The topics are denoted as T_k, with $k \in \{1 . . . K\}$. The composition of a document associated with a particular topic is referred to as topic prevalence, while the composition of words within a specific topic is known as topical content.

2.4.2.1. Effects of covariates

As mentioned earlier, the model allows for the inclusion of covariates that can influence both topic prevalence and topical content. Prevalence covariates influence the probability of observing certain topics based on the values of the covariates. On the other hand, content covariates impact the likelihood of specific words occurring within each topic. Both types of covariates are considered to account for different factors that can affect the composition of topics.

During the estimation of the Structural Topic Model (STM) for our specific case study, we incorporated prevalence covariates and content covariates. In our case study, for example, “date”, “like count”, “retweet count”, and “official” are considered as the covariates that affect the topic prevalence. While “official” is considered as content covariate, which is the author category ("Official Samsung account"=TRUE Vs "other"=FALSE) and affect the topic content.

This allowed us to explore how the source of the tweets influenced the prevalence of certain topics. Additionally, content covariates were considered to assess how specific words were associated with different topics.

2.4.2.2. Estimate the proportion

Our primary objective is to estimate the proportion of each topic within a document and then aggregate these estimates to obtain measures of topic proportions for the entire sample. The topic proportion for each document is defined as θ_d, where d ∈ {1, . . . D}. The Structural Topic Model (STM) is a generative model that accounts for the relative frequency of occurrence of words. This accounts for the possibility of documents having different lengths and counts but the same relative frequency. Given the described components, a data generative process is defined for each document, and the parameters of the model are estimated using the data sample. The distributions for each component are as follows:

$$\gamma_k \sim \text{Normal}_P\,(0, \sigma_k^2\, I_p), \quad \text{for } k = 1 \ldots K - 1$$

$$\theta_d \sim \text{LogisticNorma}_{K-1}\,(\,\Gamma'\, X_d'\,, \Sigma\,)$$

$$Z_{d,n} \sim \text{Multinomial}_K\,(\theta_d) \text{ for } n = 1 \ldots N_d$$

$$W_{d,n} \sim \text{Multinomial}_V\,(\text{B}z_{d,n}) \text{ for } n = 1 \ldots N_d$$

The process of generating D documents using a Structural Topic Model (STM) with a vocabulary size of V and K topics can be summarized as follows:

1. Draw the document-level attention to each topic, i.e. its document-level propensity θ_d from a logistic-normal generalized linear model based on a vector of document covariates X_d, containing the p covariates of document d,γ is a p x K-1 matrix of coefficients for the topic proportion and Σ is K-1 x K-1 topic covariance matrix.

2. For each word in the document, denoted by n (where n ranges from 1 to Nd):

(a) Select the topic assignment $z_{d,n}$ for the word based on the document-specific distribution over topics.

(b) Given the chosen topic, generate the observed word $w_{d,n}$ from that topic.

Figure 3 provides a representation of the process we just described.

Figure 3. Description of the STM

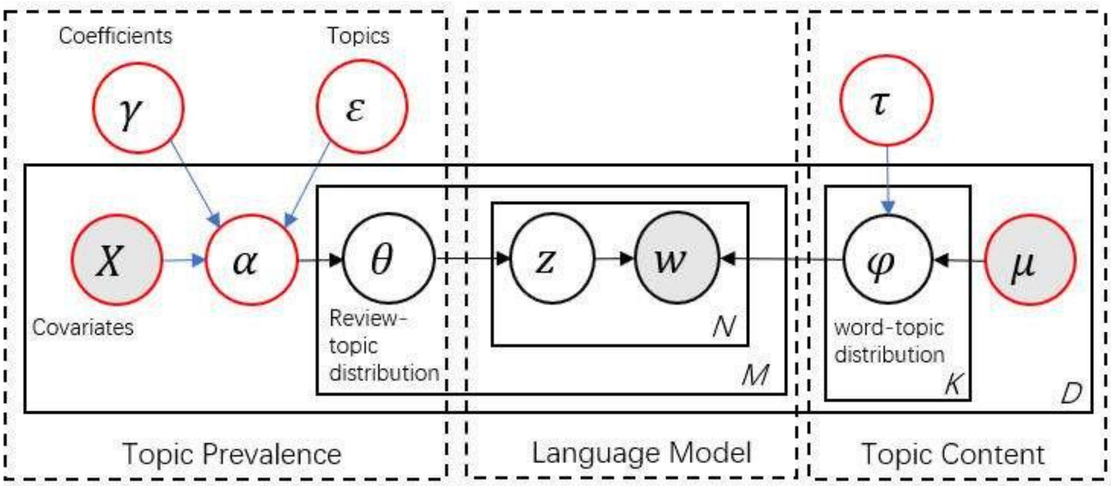
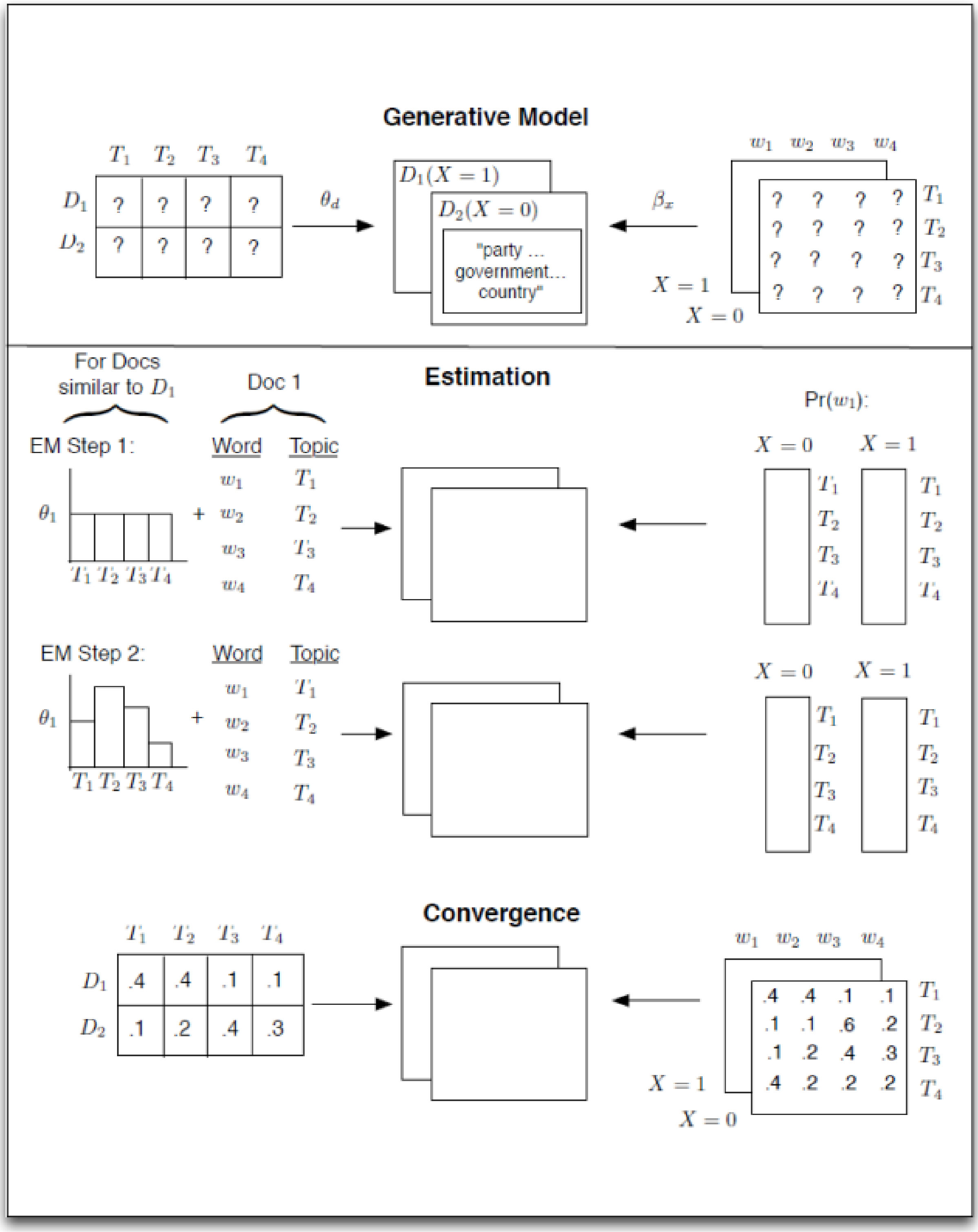


Figure 4. Heuristic description of generative process and estimation of the STM



Chapter 3

Findings and Results

3.1. Document-feature matrix (DFM) and Top features

The document-feature matrix (DFM) represents the frequency of phrases present in the studied contents (Official and other users' accounts). In this matrix, rows represent the desired documents, and columns are created based on the obtained features.

It should be noted that before creating the final DFM matrix, words defined as stopwords for the texts have been removed from the documents. Additionally, words that have been used very rarely in a very small number of texts have been identified and excluded from the document set. Moreover, words consisting of a single character (scattered letters occasionally present in the texts) or numbers have also been removed from the documents. For example, words and characters like "via", "my", "s", "l", "can", "|", "m", "i", "d", "ve", "go", "even", "also", "galaxys21", "galaxys" have been removed.

Finally, the process of creating a document-feature matrix has been performed for all the contents related to both categories. This allows for the comparison of identified terms within each category.

Table 1 presents the top features of content or most frequent words and bigrams that are more likely (higher probability of being observed) for both official and other users' accounts. The features such as "win," "get," "chance," and "devices" indicate that official Samsung accounts focus on promoting opportunities to win or acquire Samsung Galaxy S21 devices. This suggests that the product identity communicated in the advertisement campaign emphasizes the desirability and exclusivity of owning a Galaxy S21.

In the other users' accounts, the feature "Samsung" is the most prominent which means users speak more about the Samsung, followed by "video," "find," and "fast." This strongly suggests a connection to Samsung's advertisement campaign highlighting the creating Video.

This suggests that users frequently discuss or share content related to Samsung, including information about finding or experiencing the product.

Table 1. Top Features of Content: Official Samsung Accounts vs. Other Accounts

	Top features of official Samsung accounts	Count	Top features of other Users' accounts	Count
1	win	5387	Samsung	31989
2	get	4316	video	31501
3	Samsung unpacked	3871	find	25414
4	chance	3817	fast	23721
5	devices	3731	official	22789
6	re	3727	play	22742
7	need	3723	t-mobile	22513
8	click	3716	tmobiletuesdays	22452
9	signed	3706	interactive	22420
10	next	3696	epic moments	20379
11	giving	3686	ultra	19791
12	reply	3686	galaxy samsungmobilesa	15069
13	stop	3684	artscase	14498
14	tweet	3681	iPhone	12709
15	follow	3679	epic	10199

16	dm	3677	camera	10062
17	instructions	3675	love	9470
18	steps	3674	win	9413
19	opt	3674	capture	8119
20	apply	1736	get	7752

Word clouds provide a quick overview and snapshot of the main themes and concepts associated with each topic in the model.

Figure 5 and 6 showcases the word clouds, representing the most frequent used words in the other users' accounts and official accounts, respectively. It visually displays the words that have the highest probability of occurring in each category. The size and prominence of each word in the cloud are proportional to its frequency or importance within the data group. In the other users wordcloud, the words "Video", "find", "official", "fast" and "iPhone" are prominent. Whereas in the official accounts wordcloud, the words "win", "get", "next", "apply" are outstansing. The notable occurrence of words like "iPhone," "apple," and "apple_samsung" indicates that users are actively comparing the Galaxy S21 with Apple's iPhone, signifying a discernible interest in evaluating both devices.

Figure 5. Wordcloud for the users' accounts

Figure 6. Wordcloud for the official accounts

3.2. Top topics

In the next stage, structural topic modeling was performed to identify topics' top contents for both data categories. The top 10 topics for each category (Samsung and its users) along with the identified phrases for each topic have been presented in Tables 2 and 3, respectively.

To facilitate a comprehensive analysis and highlight significant differences in the choice of words and bigrams between official accounts and other users' accounts, we extracted the most frequent words and bigrams from each topic for both categories. These findings are summarized in Table 4. It showcases the specific content priorities and interests expressed by the two groups by examining the interesting and relevant differences between the words used by official accounts and other users accounts, we can gain valuable insights into the communication strategies employed by official accounts compared to user-generated content.

Table 2. Top topics in official accounts

Topics	Top words and bigrams
1	night, photos, check, mode, bts, one, never, capture, night_mode, want, time, portrait, perfect, miss, plus, just, learn, don, video, stars
2	galaxy, epic, series, get, now, samsung, everyday, learn, ready, new, experience, pre order, plus, watch, everyday_epic, check, buy_now, buy, free, ultra
3	epic, mm, us, withgalaxy, win, moments, feature, epic_moments, chance, chance_win, good, tell, epic moments, stand, fave, thing, share, miss, make, series
4	best, plus, mm, showmethegalaxys, find, share, luck, stand, now, fam, rt, like, comment, post, latest, video, hi, using, device, fast
5	ultra, withgalaxy, camera, learn, video, zoom, take, capture, epic, snap, mp, using, plus, shot, samsung, captured, design, know, every, life
6	epic, new, series, worth, every, buy, way, bts, buy_new, designed, galaxyxbts, epic_every, designed_epic, every_way, exclusive, get, get_exclusive, ultra, last, bts_twt

7	samsungunpacked, one, epic, series, one_epic, ultra, view, director, waiting, check, ultra_samsungunpacked, samsung, director_view, away, pro, walk, two, like, get, like_pro
8	win, re, devices, need, click, get, samsungunpacked, next, chance, chance_win, giving, reply, stop, signed, samsungunpacked_re, follow, tweet, dm, giving_chance, click_tweet
9	enjoy, now, get, worth, gifts, gifts_worth, buy, epic, plus, seamless, offers, exclusive, one, something, ultra, loved, season, want, youtube, premium
10	get, see, ts, right, deals, galaxybudspro, galaxysmarttag, cs, gift, apply, cs_apply, ts_cs, phone, everything, love, makes, like, link, work, nothing

Table 3. Top topics in other users' accounts

Topics	Top words and bigrams
1	video, samsungindia, samsung, contestalert, night, snap, mode, view, feature, video_snap, night_mode, resolution, get, director, director_view, samsung_samsungindia, samsung_contestalert, plus, contestalert_samsung, uncovertheepic
2	artscase, samsung, iphone, apple, art, galaxy, apple_samsung, phonecases, iphonecases, art_phonecases, artscase_art, artscase_artsmakers, artsmakers, artsmakers_artscase, iphonecases_iphone, phonecases_iphonecases, android, google, pro, iphone_pro
3	epic, epicmoments, samsungmobilesa, capture, videos, moments, create, love, take, pictures, moment, epicmoment, nationalepicday, video, memories, family, epic_moment, epic_moments, definitely, every
4	useek, fast, video, play, official, find, t-mobile, tmobiletuesdays, tmobiletuesdays_useek, interactive, video_find, interactive_video, play_interactive, useek_official, fast_play, useek_useek, useek_vía, vía, vía_useek, nobodyislistening
5	samsungmobilesa, mp, camera, feature, showmethegalaxys, photos, battery, life, within, capture, pinch, expressoshow, favourite, fav, favorite, ecr, detail, mah, details, favorite_feature
6	new, display, live, get, bts, win, epic, experience, youtube, us, event, samsungpakistan, twt, bts_twt, tgfamily, galaxyxbts, today, amazing, technicalguruji, hz
7	samsung, galaxy, ultra, samsungunpacked, samsung_galaxy, series, plus, now, get, phantom, galaxy_ultra, pro, samsunggalaxys, new, black, order, smartphone, free,

	buds, gb
8	win, tag, tell, samsungmobilesa, tell_samsungmobilesa, smart, find, smart_tag, buds, chose, apply, pro, buds_pro, ts, cs, ts_cs, cs_apply, re, draw, tag_buds
9	ultra, special, season, hoping, holiday, ultra_special, holiday_season, special_holiday, hoping_ultra, galaxyholidayhint, one, download, brought, captured, flash, unlocked, shot, one_epic, epic_shot, life
10	phone, camera, best, like, one, amazing, love, quality, just, features, time, zoom, need, now, perfect, good, content, want, super, better

Next, we will delve into a comparative analysis of the content within each topic, focusing on their respective categories. We will explore the unique characteristics of each topic within each category separately, and then compare them between the two categories.

3.2.1. Comparative analysis of topics patterns within Official and all other accounts

Upon reviewing the findings presented in Table 4, it is evident that topics 1, 4, and 5 in the official accounts category do not possess any top feature words. Notably, the top features "instructions," "steps," and "opt" were not identified in any of the topics within the official accounts category. Additionally, topics 2, 6, 9, and 10 primarily consist of the word "get," suggesting a focus on obtaining or acquiring something. However, topic 8 stands out with a wide range of top feature words such as "win," "devices," "click," "samsungunpacked," and more. This indicates a diverse range of content and promotional activities in this topic.

Conversely, for other users' accounts, the only top feature "galaxy samsungmobilesa" was not identified in any of the topics. However, it is worth noting that in all topics, at least two top feature words were identified, indicating distinct communication patterns among users' accounts. In this category, topic 5 primarily revolves around the word "camera" and the action of capturing, suggesting a focus on photography. Topic 8 stands out with the top feature words "find" and "win," indicating a potential interest in finding something or participating in contests. Topic 9 also shows a focus on capturing and mentions the word "ultra." Topics 2 and 7 feature the words "samsung" and "iphone," suggesting a comparison or discussion between the two brands.

Furthermore, it is noteworthy that topic 8 in the official accounts category and topic 4 in other users' accounts exhibit the highest number of top feature words among all the topics. This suggests that these topics have a richer and more diverse content compared to others. The presence of multiple top features in these topics indicates a broader range of subjects and discussions within these categories of accounts.

In the official accounts category, we observe that Topic 2 predominantly features the word "get," suggesting a focus on encouraging users to obtain something related to the Galaxy S21 model. Similarly, Topic 6 in official accounts also emphasizes the feature "get." Furthermore, Topics 5, 9, and 10 in the other users' accounts category showcase discussions centered around camera features, capturing moments, and expressing love for Samsung devices.

Overall, while official accounts seem to emphasize promotions, giveaways, and participation, other users' accounts showcase a wider variety of topics, including photography, contests, and discussions comparing Samsung and iPhone.

3.2.2. Comparative analysis of topics patterns between Official and all other accounts

The analysis of the most frequent words and bigrams for each topic within both the official accounts and other users' accounts separately, provides valuable insights. However, to gain a deeper understanding, we need to conduct a comparative analysis between the same topics in each category. This will allow us to identify any differences or similarities in the focus of discussions between the official accounts and other users' accounts.

By comparing the top features in each topic side by side, we can draw meaningful conclusions about the content and themes that are prevalent in these two categories.

In Topic 1, official accounts do not have any identified top feature words, while other users' accounts prominently mention "video" "samsung" and "get". This suggests that official accounts may not focus on specific content or engagement in this topic, while other users' accounts prioritize sharing videos related to Samsung and seeking user engagement. We will discuss it more later in the next sub-section.

Moving to Topic 2, the top feature for official accounts is "get," indicating their emphasis on encouraging users to take action. In contrast, other users' accounts mention "samsung" "iphone" and "artscase" suggesting discussions comparing Samsung with iPhone and mentioning a specific phone case brand. This difference can be attributed to the marketing objectives of official accounts to drive user actions and the organic conversations of other users.

In Topic 3, official accounts focus on "win" and "chance," likely promoting opportunities for users to win something. On the other hand, other users' accounts revolve around "video," "love," "capture," "epic," and "epicmoments" indicating a desire to share and discuss capturing memorable and exciting moments through videos. This divergence suggests that official accounts prioritize creating a sense of anticipation and participation, while other users' accounts center around expressing positive sentiments and sharing experiences.

For Topic 4, official accounts do not have any top feature words, while other users' accounts highlight "fast," "video," "play" and mention the official accounts of T-Mobile and T-Mobile Tuesdays. This difference suggests that official accounts may not have a significant presence or specific focus in this topic, while other users' accounts focus on fast-paced video content and interactions with T-Mobile. We will discuss it more later in the next sub-section. We will discuss it more later in the next sub-section.

In Topic 5, official accounts again do not have top feature words, while other users' accounts emphasize "camera" and "capture." This indicates that official accounts may not have prominent discussions related to cameras or capturing moments, while other users' accounts prioritize these aspects, potentially sharing their photography experiences or tips.We will discuss it more later in the next sub-section.

Moving to Topic 6, both official accounts and other users' accounts prioritize the word "get." Additionally, other users' accounts mention "win" and "epic," suggesting a shared focus on encouraging participation, winning opportunities, and highlighting epic experiences. This similarity may indicate a shared objective of engaging and exciting users.

In Topic 7, official accounts prominently feature "samsungunpacked" and "get," potentially promoting Samsung's unveiling events and encouraging user engagement. Other users' accounts also mention "samsung" and "ultra" in addition to "get," indicating a parallel interest in obtaining Samsung Ultra models. This similarity suggests that both official accounts and other users' accounts recognize the importance of Samsung's flagship events and devices.

In Topic 8, official accounts have a comprehensive list of top feature words such as "win," "re," "devices," "need," "click," and "get," along with various engagement-related terms like "follow" and "tweet." Other users' accounts, in contrast, mention "find" and "win." This discrepancy indicates that official accounts have a more diverse range of features, focusing on activities like winning, device-related discussions, and calls to action, while other users' accounts center their discussions around finding and winning.

For Topic 9, official accounts prioritize the word "get," while other users' accounts mention "capture" and "ultra." This suggests that both official accounts and other users' accounts recognize the importance of obtaining devices, but other users' accounts specifically highlight discussions related to capturing the Samsung Ultra models.

Lastly, in Topic 10, official accounts mention "get" and "apply," suggesting a focus on encouraging users to take action or apply for something. Other users' accounts, however, mention "camera" and "love," indicating a strong affinity for Samsung cameras. This difference showcases the distinct objectives of official accounts in driving actions and the organic conversations among other users, centered around expressing admiration for Samsung's camera capabilities.

Overall, The official accounts focus on encouraging users to "get" and "win" Samsung-related content, while other users' accounts engage in discussions about video content, device comparisons, and capturing memorable experiences with Samsung devices. Both categories share a common emphasis on user engagement and acquisition, as indicated by the presence of words like "get" and "win" in multiple topics. However, official accounts demonstrate a stronger emphasis on promotions, contests, and specific actions, while other users' accounts show more diverse discussions around video content, camera features, and experiences with Samsung devices.

Table 4. The most frequent words and bigrams in each topic cluster

Topics	The most frequent words and bigrams	
	Official accounts	**Other users account**
1	-	video, samsung,get
2	get	samsung, iPhone,artscase
3	win, chance	video, love, capture, epic, epicmoments
4	-	fast, video, play, official, find, t-mobile, tmobiletuesdays, interactive
5	-	camera, capture
6	get	win,epic,get
7	samsungunpacked, get	samsung,ultra,get
8	win, re, devices, need, click, get, samsungunpacked, next, chance, signed, giving, reply, stop, follow, tweet, dm	find, win
9	get	capture, ultra
10	get, apply	camera, love

3.2.3. Result visualizing

In this section, we employ visualizations to compare topics for both types of accounts, aiming to enhance the analysis of the findings presented in the preceding section.

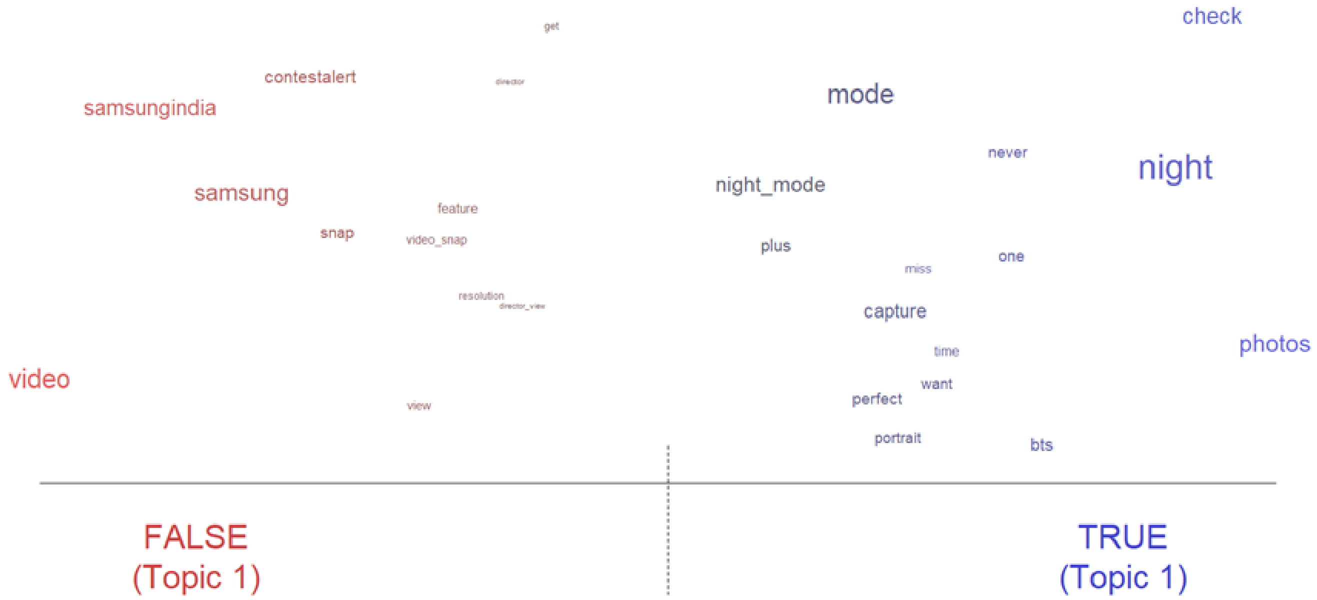
The visual plots provide valuable insights into the degree of similarity and dissimilarity between words used by officials and all other users in their tweets. A concentration of words on the left side indicates a predominance of terms from all other accounts, while a concentration on the right side signifies a prevalence of words from Official Samsung accounts ("Official Samsung accounts"=TRUE Vs "all other accounts"=FALSE). These plots offer a deeper understanding compared to the top 10 topic tables, as they showcase not only the most frequent words but also the differences in probabilities of word occurrence, which can vary in magnitude. Consequently, the plots offer a more comprehensive picture by providing additional information about the relative prominence of specific words.

In Table 4, we observed that topics 1, 4, and 5 for official accounts did not include any top feature words, making it challenging to analyze and compare these topics with their counterparts in other users' accounts. To gain a better understanding of the content within these topics, we visualize them in this section.

Figures 7, 8, and 9 provide a comparison between the two groups for the topic 1, 4, and 5, respectively. They display the words and their probabilities of occurrence in the tweets from both types of accounts. The color of the words indicates which group discusses them more frequently, with blue representing official accounts and red representing all other accounts. The size of the words reflects their general frequency of usage.

The comparison in Figure 7 illustrates the differences in topic 1 between the official accounts of Samsung (TRUE side) and all other accounts (FALSE side). On the TRUE side, which represents official accounts, we can observe that users tend to discuss characteristics like "night," "mode," "night_mode," and "photos" more frequently in their tweets. This aligns with the focus of official accounts on promoting the features of the Galaxy S21, such as its advanced camera capabilities. In fact, the new feature of the Galaxy S21, "Snap the perfect moment from video with 8K Video Snap," may contribute to the prominence of these photography-related terms in official account tweets. On the FALSE side, which represents all other accounts, users focus more on characteristics like "video" and "snap" in their tweets, potentially reflecting their interest in capturing and sharing moments from videos using the Galaxy S21's innovative capabilities. These findings provide insights into the perspectives and interests of the two account categories within topic 1, as well as how the new Galaxy S21 feature may shape the conversations among customers.

Figure 7. Comparison of topic 1 between two groups



In Figure 8, on the TRUE side, which represents official accounts, users frequently mention characteristics such as "plus," "best," "showmethegalaxys," and "share" in their tweets. These words suggest that official users focus on highlighting the premium features and qualities of Samsung devices, positioning them as superior options. Furthermore, the inclusion of "share" implies an emphasis on encouraging users to share their experiences and content related to Samsung products.

In contrast, the FALSE side, which corresponds to all other accounts, demonstrates a higher occurrence of words like "useek," "find," "play," and "video" in their tweets within topic 4. These terms indicate that general users and enthusiasts prioritize activities such as seeking information, finding resources or solutions, engaging with multimedia content, and potentially exploring various aspects of the Samsung ecosystem.

Official accounts of Samsung focus on promoting the positive aspects of their products, highlighting the "plus" features, positioning themselves as the "best" option, and engaging in campaigns like "showmethegalaxys" relating a competition among customers to create a sense of exclusivity and desirability. They also emphasize the importance of sharing experiences and content related to Samsung devices. On the other hand, all other accounts demonstrate a stronger interest in seeking information, finding resources, engaging with video content, and potentially exploring different aspects of the Samsung ecosystem.

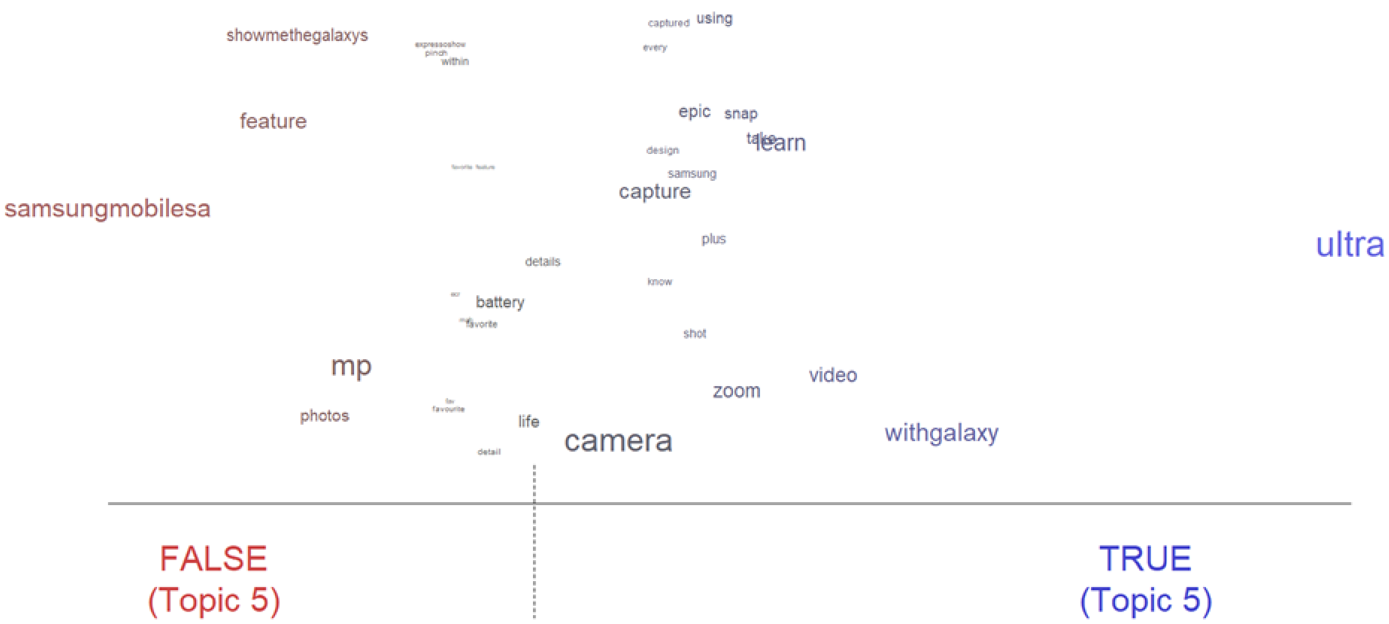
Figure 8. Comparison of topic 4 between two groups

Figure 9 showcases the word usage differences between the official accounts of Samsung (TRUE side) and all other accounts (FALSE side) within topic 5. On the TRUE side, which represents official accounts, users prominently mention characteristics such as "ultra," "camera," "video," "zoom," and "capture" in their tweets. These words indicate a strong focus on the advanced camera capabilities of Samsung devices, particularly the ability to capture high-quality photos and videos with enhanced zoom features. It suggests that official users prioritize highlighting the impressive camera performance and emphasizing the visual storytelling aspects of Samsung products.

In contrast, the FALSE side, corresponding to all other accounts, demonstrates a higher occurrence of words like "features," "showmethegalaxys," "battery," and "photos" in their tweets within topic 5. These terms suggest that general users and enthusiasts place a greater emphasis on discussing various device features, participating in Samsung-related campaigns like "showmethegalaxys," evaluating battery performance, and sharing their own photos taken with Samsung devices.

Official accounts of Samsung primarily focus on highlighting the "ultra" capabilities of the camera, emphasizing the importance of capturing high-quality "video" and utilizing advanced "zoom" features. They actively promote the ability to capture visually captivating content and engage in visual storytelling. On the other hand, all other accounts display a stronger interest in discussing various device "features" participating in Samsung campaigns like "showmethegalaxys" evaluating "battery" performance, and sharing personal "photos" captured with Samsung devices. These users appear to be more inclined towards exploring the overall device functionality, engaging in community initiatives, and assessing practical aspects such as battery life.

Figure 9. Comparison of topic 5 between two groups



3.3. The correlation network between topics

In Figure 10 we plot the topic proportions. In this case, we have a different number of topics, with contents. The expected topic proportion refers to the anticipated distribution or prevalence of topics within a given corpus of documents. It represents the probability of each topic occurring in the overall dataset.

The topics are ranked based on their highest probability of occurrence within the all analyzed dataset. Let's analyze and compare these topic proportions with the related analysis we have discussed so far.

Topic 4, which includes words like "find," "video_find," and "tmobile," seems to highlight the importance of discovering and accessing content related to the Galaxy S21. This aligns with our previous observations regarding the emphasis on finding and playing videos among all other user accounts. While topic 3, characterized by words such as "remember," "include," and "epic_moments," suggests a focus on capturing and commemorating memorable experiences. This is consistent with our earlier analysis, where we identified a strong association between other user accounts and terms like "video," "capture," and "epicmoments." However, topic 7, represented by terms like "ultra_samsungunpacked," "samsungevent," and "unpacked," likely pertains to official Samsung accounts discussing the Samsung Unpacked event and promoting the Ultra variant of the Galaxy S21. This aligns with our previous findings of official accounts being associated with terms like "samsungunpacked" and "ultra."

Topic 10, featuring words like "easier," "yet" and "nothing," could indicate discussions around the ease of use and user experience of the Galaxy S21. Although this topic was not specifically highlighted in our previous analyses, it suggests that user accounts may have mentioned the simplicity or lack of complications in relation to the device. In addition, topic 8, with terms like "entered_draw," "entered," and "pro_ts," might indicate engagement with contests or promotional activities related to the Galaxy S21. This corresponds to our earlier identification of official accounts frequently using words like "win," "chance," and "apply."

The remaining topics also provide different aspects of the Galaxy S21, such as portrait mode, camera capabilities, and seasonal themes.

Figure 10. Topic prevalence for the contents

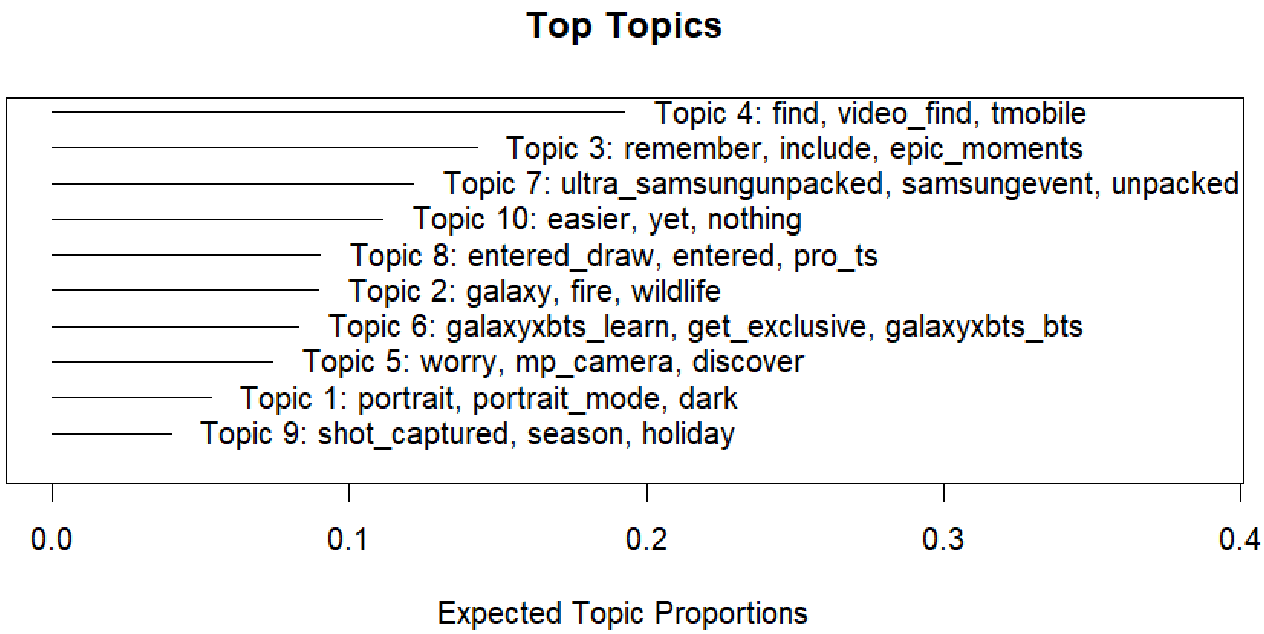
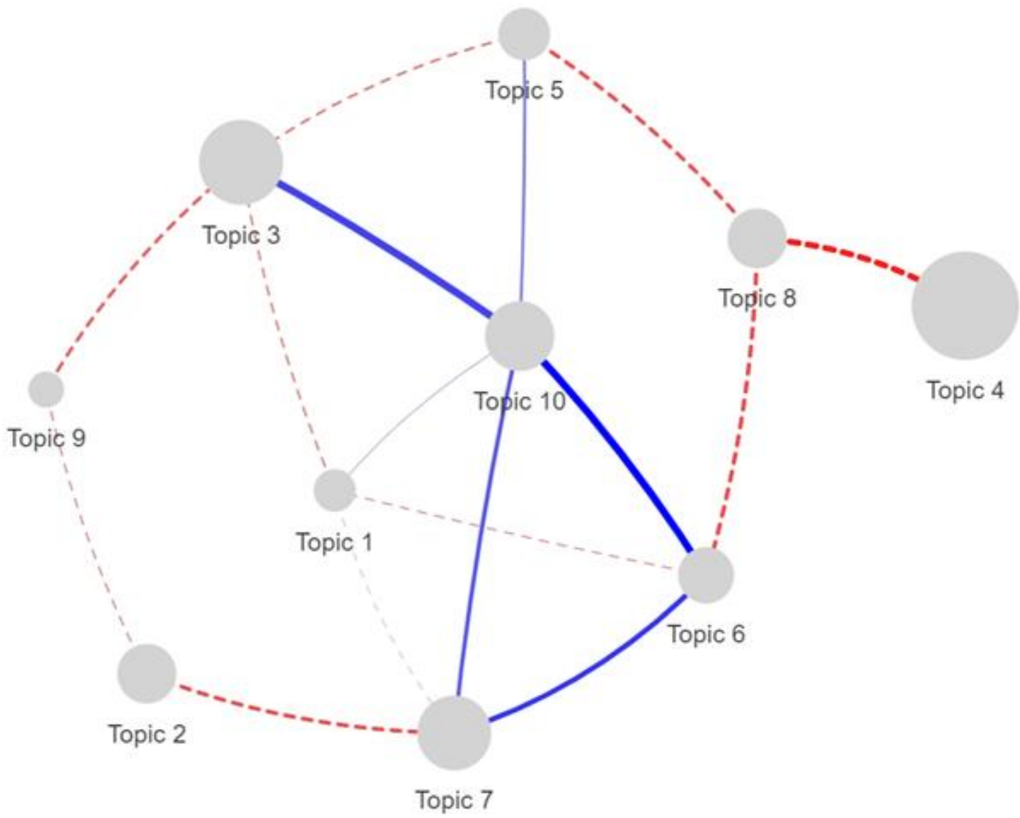


Figure 11 shows correlations between topics for 10 topics, providing information about the topics that are more likely to be observed together. Topics correlation refers to the degree of association or similarity between different topics. It measures how frequently topics co-occur or are discussed together in the analyzed text data. A positive topic correlation (positively correlated topics are connected by blue edges) suggests that the topics are often mentioned together, indicating a potential thematic relationship or overlap in the content. On the other hand, a negative topics correlation (negatively correlated topics are connected by red edges) indicates that the topics are less likely to co-occur, suggesting distinct themes or separate discussions. The size of the nodes is proportional to the number of words in our corpus devoted to each topic (see Figure 10).

First, we observe that Topics 3, 5, 6, and 7 show a positive correlation with Topic 10. These topics are characterized by words such as "remember," "include," "epic_moments," "worry," "mp_camera," "discover," "galaxyxbts_learn," "get_exclusive," and "galaxybts_bts." The positive correlation suggests that discussions related to remembering epic moments, camera features, and exclusive content surrounding Samsung galaxyS21 are often connected. This indicates a cohesive narrative where users are engaged in sharing experiences, discovering features, and participating in exclusive content related to the device.

On the other hand, Topics 5, 4, and 6 exhibit a negative correlation with Topic 8. These topics are characterized by words such as "worry," "mp_camera," "discover," "find," "video_find," "tmobile," and "galaxyxbts_learn." The negative correlation suggests that discussions revolving around users' concerns, camera features, and discovering new aspects of Samsung galaxyS21 do not align closely with the content associated with entering draws or participating in professional activities (Topic 8). It appears that these topics represent different discussions or aspects related to the device that is distinct from the content of Topic 8.

Figure 11. Topics correlation



Discussion

In this section, we discuss the findings and results of our research, addressing research questions and providing relevant insights. By examining the document-feature matrix, word clouds, top topics for both official and other users' accounts and, topics correlation, we can draw conclusions regarding the communication strategies employed by Samsung Mobile and the discussions among users.

The findings reveal that the product identity conveyed in the campaigns aligns moderately with user-generated content. Both Samsung's official campaigns and user discussions emphasize the high-quality camera, powerful performance, and sleek design of the Galaxy S21. However, Figure 5 highlights additional features and aspects that users frequently mentioned, such as battery life and durability, which were not as prominently addressed in Samsung's campaigns. This indicates the need for Samsung to further align their messaging with the preferences and perceptions expressed by users.

Given Table 4, the analysis indicates a successful online communication strategy. The sentiment expressed by users in the comments is predominantly positive, reflecting satisfaction with the Galaxy S21 and indicating a favorable response to Samsung's online communication efforts.

The analysis indicates that users align primarily with product features, design, and user experience, which are consistent with Samsung's intended messaging. Users appreciate the innovative features and sleek design of the Galaxy S21, as well as the overall usability and performance of the device.

However, users also bring their individual perspectives and preferences into the discussions. In addition, users also tend to make comparisons between Samsung and iPhone, probably about characteristics such as the camera. They often share personal stories, opinions, and experiences related to the Galaxy S21, which may deviate from Samsung's intended messaging.

Overall, the findings suggest a substantial level of alignment between Samsung Mobile's advertisement campaigns and user perceptions, highlighting the success of their online communication strategies. However, there are opportunities for further improvement, such as aligning messaging strategies with user preferences and incorporating user feedback to enhance the overall effectiveness of the campaigns.

Conclusions

The aim of this study is to investigate the alignment between the content used in the advertising campaigns offered by Samsung Mobile on Twitter and what users communicate about the product in their tweets. To this purpose, we applied STM to examine Samsung Mobile's campaigns on Twitter and made comparison between the tweets by Samsung official accounts and all other users accounts.

In conclusion, the findings suggest that there is alignment between the product identity communicated in Samsung Mobile's advertisement campaigns and what users communicate about the product. However, there are some differences in communication content priorities between official accounts and other users' accounts, indicating a diverse range of discussions and interests among users. These insights can help Samsung Mobile refine its communication strategies and better understand the preferences and interests of its target audience.

These findings have implications for both academic research and marketing practitioners, emphasizing the importance of aligning advertisement campaigns with user communication to ensure a consistent and impactful brand identity. Further research could delve deeper into specific advertisement elements, target audience segments, and other social media platforms to gain a comprehensive understanding of the alignment between product identity and user communication using STM analysis.

List of Figures

List of Tables

List of References

Adjeng Mariana Febrianti, R., Kusriyantini, D., Safitri, F. A., & Roespinoedji, D. (2022). INCREASING BRAND LOYALTY THROUGH SOCIAL MEDIA MARKETING AND BRAND AWARENESS (SAMSUNG CASE STUDY). Central Asia & the Caucasus (14046091), 23(1).

Arsalani, M., Esmaeilkhoo, H., & Taghipour, M. (2020). Investigating the Effect of Social Media Marketing Activities on Brand Awareness. Management, 4(2), 18-31.

Bendle, N., Farris, P., Pfeifer, P., & Reibstein, D. (2010). Metrics that matter–to marketing managers. Marketing: Journal of Research and Management, 6(1), 18-23.

Braverman, S. (2015). Global review of data-driven marketing and advertising. Journal of Direct, Data and Digital Marketing Practice, 16(3), 181-183.

Cassone, F. (2020). Using sentiment analysis to measure the outcome of marketing campaigns on Twitter, an empirical approach [Master's thesis, LUISS Guido Carli]

Chaney, P. (2009). The digital handshake: Seven proven strategies to grow your business using social media. John Wiley & Sons.

Charlesworth, A. (2014). Digital marketing: A practical approach. Routledge.

Davenport, T. H. (2013). Enterprise analytics: Optimize performance, process, and decisions through big data. Pearson Education.

Dremel, C., Herterich, M. M., Wulf, J., & Vom Brocke, J. (2020). Actualizing big data analytics affordances: A revelatory case study. Information & Management, 57(1), 103121.

Ektoros, E., Gregoriades, A., & Georgiades, M. (2019). Online Consumers' Opinions Analysis for Marketing Strategy Evaluation. In ICEIS (1) (pp. 266-273).

Fulgoni, G. M., Pettit, R., & Lipsman, A. (2017). Measuring the effectiveness of branded content across television and digital platforms: How to align with traditional marketing metrics while capturing what makes branded content unique. Journal of Advertising Research, 57(4), 362-367.

Fulgoni, G., & Lipsman, A. (2014). Digital game changers:: how social media will help usher in the era of mobile and multi-platform campaign-effectiveness measurement. Journal of Advertising Research, 54(1), 11-16.

Gaitniece, E. (2018, September). Digital marketing performance evaluation methods. In CBU International Conference Proceedings (Vol. 6, pp. 135-140).

Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. Journal of interactive marketing, 27(4), 242-256.

Goodwin, T. (1999). Measuring the effectiveness of online marketing. Market Research Society. Journal., 41(4), 1-6.

Gregoriades, A., Pampaka, M., Herodotou, H., & Christodoulou, E. (2021). Supporting digital content marketing and messaging through topic modelling and decision trees. Expert systems with applications, 184, 115546.

Gutnik, S. (2021). Application of Data Mining and Machine Learning Methods to Enhance the Effectiveness of Digital Marketing Strategies. In Digital strategies in a global market (pp. 131-144). Palgrave Macmillan, Cham.

Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We’re all connected: The power of the social media ecosystem. Business horizons, 54(3), 265-273.

He, L., Han, D., Zhou, X., & Qu, Z. (2020). The voice of drug consumers: online textual review analysis using structural topic model. International Journal of Environmental Research and Public Health, 17(10), 3648.

Jamali, M., & Khan, R. (2018). The impact of consumer interaction on social media on brand awareness and purchase intention! Case study of Samsung. Journal of Marketing, 114.

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. Journal of marketing, 57(1), 1-22.

Kelly, N. (2012). How to measure social media: A step-by-step guide to developing and assessing social media ROI. Que Publishing.

Ketter, E., & Avraham, E. (2012). The social revolution of place marketing: The growing power of users in social media campaigns. Place Branding and Public Diplomacy, 8(4), 285-294.

Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. Journal of marketing, 80(1), 7-25.

Laudon, K. C., & Traver, C. G. (2016). Business Technology Society.

Li, Z., Fan, Y., Jiang, B., Lei, T., & Liu, W. (2019). A survey on sentiment analysis and opinion mining for social multimedia. Multimedia Tools and Applications, 78(6), 6939-6967.

Lumb, C. (2014). The Customer Decision Process and User Interaction in E-commerce. The University of Leeds.

Nayab, G., Bilal, M., & Shrafat, A. (2016). A brand is no longer what we tell the customer it is-it is what customers tell each other it is:(Lahome). Sci. Int, 28, 2725-2729.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, WooK, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rathore, A. K., & Ilavarasan, P. V. (2020). Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products. International Journal of Information Management, 50, 111-127.

Reisenbichler, M., & Reutterer, T. (2019). Topic modeling in marketing: recent advances and research opportunities. Journal of Business Economics, 89(3), 327-356.

Rosário, A., Moniz, L. B., & Cruz, R. (2021). Data science applied to marketing. Journal of Information Science and Engineering, 37(5), 1067-1081.

Rosen, D. E., Schroeder, J. E., & Purinton, E. F. (1998). Marketing high tech products: lessons in customer focus from the marketplace. Academy of Marketing Science Review, 6, 1-17.

Santagiustina, C. R. M. A., & Warglien, M. (2022). The architecture of partisan debates: The online controversy on the no-deal Brexit. Plos one, 17(6), e0270236.

Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge, 6(2), 92-102.

Schivinski, B., & Dabrowski, D. (2016). The effect of social media communication on consumer perceptions of brands. Journal of Marketing Communications, 22(2), 189-214.

Sierra, C. G. (2020, May). Impact through digital marketing tools in social networks and web analytics in higher education institutions. In IOP Conference Series: Materials Science and Engineering (Vol. 844, No. 1, p. 012050). IOP Publishing.

Singh, M., & Singh, G. (2018). Impact of social media on e-commerce. International Journal of Engineering & Technology, 7(2.30), 21-26.

Singh, S., Sao, A., Nagare, T. B., & Dharmarajan, A. (2017). Role of social media marketing in brand building: the new age marketing strategy. International journal of scientific research, 5(9).

Song, H., Wang, J., & Han, H. (2019). Effect of image, satisfaction, trust, love, and respect on loyalty formation for name-brand coffee shops. International Journal of Hospitality Management, 79, 50-59.

Spiller, L., & Tuten, T. (2015). Integrating metrics across the marketing curriculum: The digital and social media opportunity. Journal of Marketing Education, 37(2), 114-126.

Strauss, J., Frost, R., & Sinha, N. (2014). E-marketing (p. 496). Upper Saddle River, NJ: Pearson.

Sundsøy, P., Bjelland, J., Iqbal, A. M., Pentland, A., & Montjoye, Y. A. D. (2014, April). Big data-driven marketing: how machine learning outperforms marketers' gut-feeling. In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (pp. 367-374). Springer, Cham.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. Decision sciences, 22(2), 219-240.

Wang, P., Liu, Y., Meytlis, M., Tsao, H. Y., Yang, J., & Huang, P. (2014, February). An efficient framework for online advertising effectiveness measurement and comparison. In Proceedings of the 7th ACM international conference on Web search and data mining (pp. 163-172).

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. Journal of Marketing, 80(6), 97-121.

Wheeler, A. (2014). Designing brand identity: An essential guide for the whole branding team. J. Brand Strategy, 2, 416-417.

Vu, T. P. Q., & Vu, T. T. T. (2021). The Impact of Social Media Marketing on Brand Loyalty: Case Study of Samsung Smartphones in Vietnam.

Thinh, T. T. X. (2021). Textual Analysis of Content Strategy in Facebook Fanpage of Samsung Galaxy A Series during Its Launched Campaigns in Vietnam.

Järvinen, J., & Karjaluoto, H. (2015). The use of Web analytics for digital marketing performance measurement. Industrial Marketing Management, 50, 117-127.

Ibrahim, N. F., & Wang, X. (2019). Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media. Computers in Human Behavior, 96, 32-45.

Barry, A. E., Valdez, D., Padon, A. A., & Russell, A. M. (2018). Alcohol advertising on Twitter—a topic model. American Journal of Health Education, 49(4), 256-263.