Instituto Superior
de Contabilidade
e Administração
Politécnico de Coimbra

Business Intelligence Applied to Sentiment Analysis in a HIE

Bruno dos Santos Varela

ISCAC | 2021

Bruno dos Santos Varela

Business Intelligence Applied to Sentiment Analysis in a Higher Education Institution

Coimbra, October 2021

Instituto Superior de Contabilidade e Administração

Politécnico de Coimbra

Bruno dos Santos Varela

Business Intelligence Applied to Sentiment Analysis in a Higher Education Institution

Dissertation submitted to the Higher Institute of Accounting and Administration of Coimbra to fulfil the requirements for obtaining a **master's degree in Management Information Systems**, conducted under the guidance of Professor Isabel Maria Mendes Pedrosa and co-supervised by Professor Jorge Fernandes Rodrigues Bernardino.

Coimbra, October 2021

TERMO DE RESPONSABILIDADE

I declare that I am the author of this Dissertation, which is an original and unpublished work, which has never been submitted to another Higher Education Institution to obtain an academic degree or other qualification. I further certify that all citations are properly identified and that I am aware that plagiarism is a serious lack of ethics, which could result in the annulment of this dissertation.

Acknowledgements

Finishing this important chapter of my personal and professional life would not be possible without the help and understanding of those around me.

To my professors a big thank you. To Professor Isabel Pedrosa for her problem-solving skills and comprehension. To Professor Jorge for his great knowledge in scientific writing and generosity in sharing it. I will remember all those late Monday meetings after work.

A big thank you to Dra. Filipa Freitas for allowing me access the data without any constrains.

To my mother for all the support.

To my sister, Brena, to whom I dedicate this work.

ABSTRACT

Social media allows institutions to not only publicize their work and get feedback from the community about it, but also to keep in touch with their alumni network and foster conversations between the academic community. While sentiment analysis allows a better understanding of what is being said about a brand and how to improve the use of this communication platform. The main goal of the current work is to build a Business Intelligence System for a Higher Education Institution (HEI) based on content extracted from social media. So, Posts, likes, dislikes, shares, comments and number of visits were extracted from Facebook, Google Maps Reviews, Instagram, LinkedIn, Student Forums, Twitter and YouTube. With this data and the ETL process a Data Warehouse (DW) in *SQL Server* and 17 Dashboards in *Power BI* were developed. Posts that had the most likes were about reporting a death of someone from the school, the school mascot, the pandemic or welcoming new students. Overall, the weekends were the days with more interactions. Students are concerned about accommodation, transport, and the school academic offer. This analysis allows a better understanding of what is being said about this HEI and how to improve the communication strategy.

Keywords: Business Intelligence, Sentiment Analyses, Social Media, Digital Marketing, Text Mining, Higher Education Institutions, Information Behaviour, Social Data

RESUMO

As redes sociais permitem que as instituições não apenas divulguem seu trabalho e recebam feedback da comunidade sobre ele, mas também mantenham contato com sua rede de ex-alunos e fomentem o diálogo entre a comunidade acadêmica. Já a análise de sentimentos permite um melhor entendimento do que se fala sobre uma marca e como melhorar o a sua estratégia de comunicação. O principal objetivo do presente trabalho é construir um Sistema de Business Intelligence para uma Instituição de Ensino Superior (IES) a partir de conteúdos extraídos das redes sociais. Assim, posts, gostos, não gostos, partilhas, comentários e número de visitas foram extraídos de Facebook, Google Maps Reviews, Instagram, LinkedIn, Forums de estudantes, Twitter e YouTube. Com esses dados e através do processo de ETL foram desenvolvidos uma Data Warehouse (DW) em *SQL Server* e 17 Dashboards em *Power BI*. Os posts com maior número de gostos foram sobre como sobre a morte de alguém da escola, a mascote da escola, a pandemia ou o acolhimento de novos alunos. No geral, os finais de semana foram os dias com mais interações. Os alunos estão preocupados com a acomodação, transporte e a oferta acadêmica da escola. Esta análise permite um melhor entendimento do que se fala sobre esta IES e como melhorar a utilização desta plataforma de comunicação.

Palavras-chave: *Business Intelligence*, Análise de Sentimentos, Redes Sociais, Marketing Digital, *Text Mining*, Instituições de Ensino Superior, Comportamento Informacional.

RÉSUMÉ

Les médias sociaux permettent aux établissements non seulement de faire connaître leur travail et d'obtenir des commentaires de la communauté à ce sujet, mais aussi de rester en contact avec leur réseau d'anciens élèves et de favoriser les conversations entre la communauté universitaire. Tandis que l'analyse des sentiments permet de mieux comprendre ce qui se dit sur une marque et comment améliorer l'utilisation de cette plateforme de communication. L'objectif principal du travail actuel est de construire un système d'intelligence d'affaires pour un Etablissement d'Enseignement Supérieur (EES) basé sur le contenu extrait des médias sociaux. Ainsi, les publications, les goûts, les dégoûts, les partages, les commentaires et le nombre de visites ont été extraits de Facebook, Google Maps Reviews, Instagram, LinkedIn, Student Forums, Twitter et YouTube. Avec ces données et le processus ETL, un Entrepôt de Données (ED) dans *SQL Server* et 17 tableaux de bord dans *Power BI* ont été développés. Les publications qui ont eu le plus goûts concernaient le signalement du décès d'un membre de l'école, la mascotte de l'école, la pandémie ou l'accueil de nouveaux élèves. Dans l'ensemble, les week-ends étaient les jours avec plus d'interactions. Les étudiants sont préoccupés par l'hébergement, les transports et l'offre académique de l'école. Cette analyse permet de mieux comprendre ce qui se dit sur cet EES et comment améliorer l'utilisation de cette plateforme de communication.

Mots-Clés : Business Intelligence, Analyse des Sentiments, Médias Sociaux, Marketing Numérique, Text Mining, Etablissements d'enseignement Supérieur, Comportement Informationnel

INDEX

FIGURES INDEX

TABLES INDEX

List of abbreviations

API	-	*Application Programming Interface*
CRISP-DM	-	*Cross-industry standard process for data mining*
CSV	-	*Comma-separated values*
DAX	-	*Data Analysis Expressions*
DBMS`	-	*Database Management Systems*
DM	-	*Digital Marketing*
DW	-	*Data Warehouse*
HEI	-	*High Education Institution*
MBA	-	*Master of Business Administration degree*
PHP	-	*Hypertext Preprocessor*
SQL	-	*Structured Query Language*
URL	-	*Uniform Resource Locator*

INTRODUCTION

Digital Platforms can have a positive or a negative impact on brands. Knowing what sentiments users have towards a brand allows companies to improve their service and receive feedback without the need of satisfaction questionnaires. On the other hand, having a deep understanding of Digital platforms can significantly benefit business.

Sentiment analyses and social media are two connected areas, and their study can benefit companies. In the case of HEIs social media allows institutions to not only publicize their work and get feedback from the community about it, but also to keep in touch with their alumni network and foster conversations between the academic community. While sentiment analysis allows a better understanding of what is being said about this HEI and how to improve the use of this communication platform.

Very few papers are made about Sentiment Analysis in Portuguese Higher Education Institutions. In C. L. Santos et al. (2018) analyses reviews made by former international students about their HEI. This analysis is made through the application of the Biterm Topic Model and 1925 reviews were extracted from the platform *iagora.com*. This method collects all unordered word-pairs (biterms) from each short text or a fixed-length text window of normal texts, considering the whole corpus as a mixture of topics (Zhu et al., 2018). It was shown that HEIs who "*financially support students cost of living, provide courses in English, and promote an international environment*" (Santos et al., 2018) were the most attractive ones to foreign students. Although focusing on a specific group of students, this paper shows how the application of sentiment analyses and Text Mining can benefit HEIs.

In Troisi et al. (2018) 12 million posts workwise from 2017, having as sources "Twitter, Instagram, blogs, forums, Google+, online newspapers, Pinterest, YouTube" were extracted. The keyword was "University", and the goal was to find what factors affected students when choosing a university. As findings, students have shown little interest for topics like scientific research, international mobility, external financing and social influence. On the other hand, the topics that interested students the most were "training offer, followed by physical structure, work opportunities, prestige, affordability, communication, organisation, environmental sustainability"(Troisi et al., 2018). This paper had a large sample and because of that it is difficult to fully represent the difference of cultures among students. By choosing only a single keyword in English many words

related to the topic were excluded. Even with the large sample constrain, findings are reliable since many studies had the same results.

Other authors choose a different approach, considering feelings related to a certain brand. For instance, in Xu et al., (2016) a prediction model to measure brand personality was developed making a survey to brands and extracting data from Twitter and Glassdoor. The study had as sources: 1996214 brand followers' descriptions and tweets, 312400 employee reviews from Glassdoor, 680056 tweets made from brands and a survey to 219 brands. According to Xu et al., (2016) brand personality can be described as an effect of the user imaginary which is a "*set of human characteristics associated with a brand*". In other words, the brand personality is how customers view the brand. In the study, for instance, Walt Disney is considered to be a family-oriented brand while Apple a young brand. When surveying the personality traits of a brand, the feelings that it awakens are also surveyed. Also, this personality analysis is also easy to perceive for all stakeholders. The development model allows measuring psychological characteristics.

The main goal of the current work is to build a Business Intelligence System for a HEI based on content extracted from social media. This work as also specific goals, namely:

-make a contextualization on Business Intelligence, Sentiment Analysis, Digital Marketing and Social Media;

-to extract posts that related to the HEI from more than one source of information;

-to make Dashboards with the extracted information;

-to build a Data Warehouse;

-to find strengths and points to improve in the HEI's communication strategy.

One of the methods used were documentary research for literature review, with the study of written documents disseminated with similar subjects. Also, the Extract, Transform and Load process was used to build the DW. In terms of softwares employed, *SQL Server* was used to build the DW and for the reports *Power BI* was used.

As results this work proves that Sentiment Analysis can be done to any type of company regardless of its size and financial resources. On the other hand, this work provides research in an area where there is little scientific production, namely the sentiment analysis applied to HEIs. Data were extracted from Facebook, Google Maps Reviews, Instagram, LinkedIn, Students Forums, Twitter and YouTube. A Data Warehouse were

made with the extracted data. With the data extracted 17 Dashboards were made. Based on the interpretation of the Dashboards the strengths and points to improve in the HEI's communication strategy were studied. Also, posts regarding grief (reporting a death of someone from the school), the school mascot, the pandemic or welcoming new students were the ones with more likes. Overall, the weekends were the days with more interactions, but this varies according to the social media and the month

Portugal has a high number of HEIs and is known for the good quality of its teaching. Being very usual for Portuguese HEIs to be in international rankings, for instance 12 institutions were on the *Times Higher Education's global portfolio* of university rankings for 2022 (Times Higher Education (THE), 2021). Also, the country belongs to several international communities, which facilitates the access to higher education by foreign students. For instance, the *Comunidade de Países de Língua Portuguesa* and the European Union. On the other hand, having such a large number of institutions makes the high education market competitive, so knowing the target audience and their preferences well becomes essential to attract new students. Is noteworthy that Portuguese schools are also growing in the number of vacancies available and students enrolling into Higher Education (Direção-Geral de Ensino Superior, 2021a).

This Master Dissertation is structured as follows. Firstly, the methodologies used are described in the 1st Chapter. The 2nd Chapter is formed by the literature review, analysing similar studies and related work. In this topic is discussed: Business Intelligence and Sentiment Analysis, social media and Digital Marketing. In 3rd Chapter all the steps until the construction of the DW are described according to the Extract, Transform and Load process. In this topic is discussed: the tools used, the Multidimensional Model, Data Extraction and Transformation, Data Load. In the 4th Chapter the results are shown in the form of Dashboards for each information source. In the 5th Chapter, recommendations are made based on the information presented on the Dashboards. Showing points to be improved and strengths in the HEI Communication strategy. The last Chapter presents the conclusions, limitations and ideas for future work.

1. METHODOLOGY

In this section is described the methods used and its importance for the current work. Also, a summary of Higher Education in Portugal and the Higher Education institution in this study is presented. Two methods were used documentary research for the literature review and the ETL process to build the DW.

One of the methodologies employed is documentary research. First, using written documents disseminated, namely scientific papers published in the last decade, with similar research topics, to allow the context of this thesis (Ribeiro & Malheiro Da Silva, 2009). Information can suffer different interpretations. There are models of data interpretation, one of which is the Tom Wilson model (Wilson, 1999), which suggests that the characteristics of information satisfaction are related to the multi-conditioned characteristics of the researcher, namely economic, emotional, educational, demographic, interpersonal and environmental that affect access to information. This model is summarized in Figure 1, showing that information seeking behaviour arises because of a need perceived by an information user. To satisfy this need users demand upon information sources and services which may result in finding the relevant information (Wilson, 1999). In this work the documentary research in the 1st and 2nd Chapter for the literature view. The Tom Wilson model were used to retrieve the scientific papers and to validate the extracted data. To find the scientific papers that sustains this study, searches were done in Google Scholar (https://scholar.google.com/), *Spinger Link* (https://link.springer.com/) and *Taylor & Francis Online: Peer-reviewed Journals* (tandfonline.com). Some of the Keywords were: Business Intelligence, Sentiment Analyses, Social Media, Digital Marketing, Text Mining, Higher Education Institutions, Information Behaviour, Social Data, Social Media in HEI, Social Media usage by HEI.
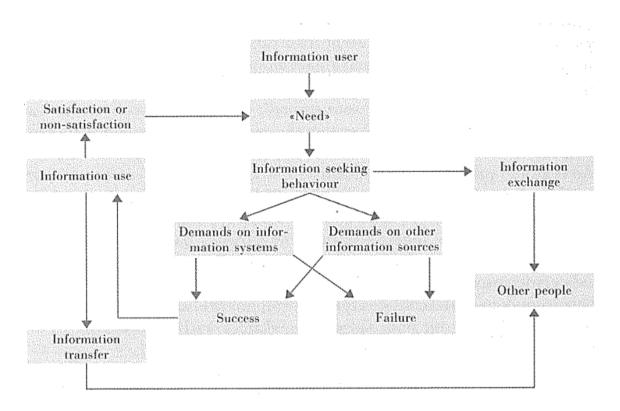
Figure 1 Wilson's model of information behaviour (Wilson, 1999)

Also, the Extract, Transform and Load (ETL) process is applied in this work. This process can be defined as "*pieces of software responsible for the extraction of data from several sources, their cleansing, customization and insertion into a DW*" (Vassiliadis et al., 2002). This process has gained increasing importance with the huge amount of data generated by billions of devices (for example, Internet of Things). The process is made from 3 steps, the extraction of data from the data source, the transformations applied to this data and the load of it in a DW. Next the steps are described in further detail. In the first step of ETL is data extraction and it can be divided into two phases. The first extraction, corresponds to the first loading of the data into the empty DW, and the subsequent incremental extractions of the data that will update the DW with the new data that they gradually emerge from the daily activity of the organization over time (Kimball & Ross, 2013). This second extraction is called the DW update. Once the data is properly extracted, follows the second step of ETL process, transform, where transformations and cleaning are done to improve data quality. It is at this stage that the anomalies commonly found in the data (duplication of data, different representations of value, data with errors, missing data) are removed, allowing the data to be presented in a homogeneous, consistent, and error-free format. In this work the Extraction and Transformation steps are merged to reduce the number of queries. After completing the entire transformation

process follows the third step, data load. In this steep data from the area of preparation is loaded into the area of presentation where they are stored in Data Marts or Cubes.

1.1   Higher Education in Portugal

One of the most widely used models in Data Science is Cross-industry standard process for data mining (CRISP-DM). This model, developed in 2000 contains six phases that describe the data science life cycle: 1.st Business Understanding; 2nd Data Understanding; 3rd Data Preparation; 4th Modelling; 5th Evaluation; 6th Deployment (Wirth & Hipp, 2000) as shown in figure 2. One of the most important phases, is business understanding, being shown that any data mining project starts with "Business Understanding" for the project to provide high value information (Huber et al., 2019). So, in the section the Higher Education in Portugal system is briefly described.
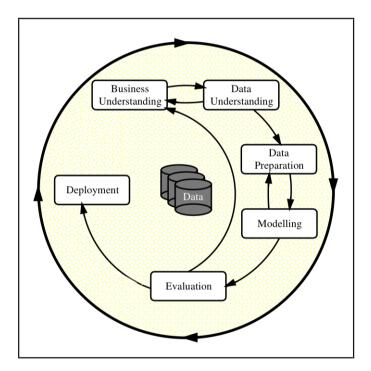
Figure 2   Phases of the Current CRISP-DM Process Model for Data Mining (Wirth & Hipp, 2000)

The Portuguese Higher Education system is based on a binary system between University education and Polytechnic education. University education is described by the Portuguese Directorate General for Higher Education (DGES) as a type of teaching aimed to "*ensure solid scientific and cultural preparation and provide technical training to enable for the exercise of professional and cultural activities*", focusing on "*innovation and critical analysis*" (Directorate General for Higher Education, 2021). While Polytechnic education is focused on "*understanding and solving concrete problems and aims at providing a*

solid cultural and technical training at the higher level", aiming for the "*exercise of professional activities*" (Directorate General for Higher Education, 2021). So, polytechnical education is considered to be more practical, while university education is more theoretical. It should be noted that a university has several faculties according to area of study and polytechnic institutions have several schools, also according to their studying nature.

Is noteworthy, the majority of HEIs in Portugal are Polytechnics. From the 106 HEI 67(63,2%) are Polytechnics and 39(36,8%) are Universities (Direção-Geral de Ensino Superior, 2021b).

1.2 The Higher Education Institution Analysed

In this section the HEI business is described, explaining its Organizational System, the school's academic offers and its presence in social media.

The HEI in analysis belongs to polytechnical education and was founded in the academic years of 1972-1973 and functions was a business school. The first course was Accounting but, in the years of 1998/1999, other courses were added. Namely: Business Management, Accounting and Audit, Management and Informatics. Recently, the school been to a rebranding process, changing its name and adding more courses. This schools offers bachelor's degrees, master's degrees, short/brief courses, and post-graduate courses.

The school belongs to a bigger entity with schools in other areas of study. This bigger institution was six other schools in the areas of agriculture, education, technology, health, and engineering. The school in study represents the seconds largest school and has about a quarter of the total of students in the polytechnical institution.

The HEI is very present in social media with an average of 1123 posts by year in the last 5 years. The figure 3 shows the number of posts by Year from 2012 to 2020 in Facebook, Instagram, Twitter and YouTube. The number of posts had increase significantly from 2015 to 2016. In 2020 the school had 699 posts.
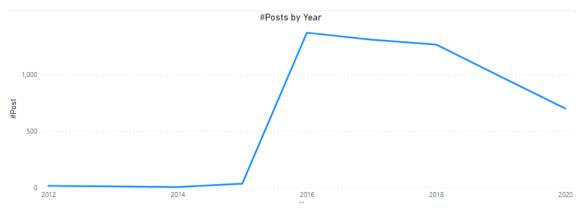
Figure 3 Number of posts by Year in Facebook, Instagram, Twitter and YouTube

So, social media has a notable importance for this school. Having an online presence al least since 2011, with the creation of the YouTube channel, as is shown on the next chapter.

2   LITERATURE REVIEW

In this Chapter contents are organized from the overall view to detail. The macro subjects of Business Intelligence and Sentiment Analysis, Digital Marketing and social media are described. In the detail part is shown how a Business Intelligence system based on social media can benefit the companies. Is given a contextual background on the social media in this study. The tendencies to social media are described. Some of the threats of social media are shown. Also, is made an analysis on how HEIs use social media. The goal of this Chapter is to contextualize the reader on these topics before describing the steps to develop the Business Intelligence System.

2.1   Business Intelligence and Sentiment Analysis

In this section the concept of Business Intelligence and Sentiment Analysis are described. Showing how the incorporation of these two areas can benefit any type of companies.

Business Intelligence systems combine data gathering, data storage, and knowledge management with analytical tools to present complex internal and competitive information to planners and decision makers (Negash & Gray, 2008), aiding in the decision making process. The figure 4 shows how Business Intelligence gathers data from different sources and though the ETL process builds a Data Warehouse. From the data in the Data Warehouse information can be displayed in the forms of Dashboards.
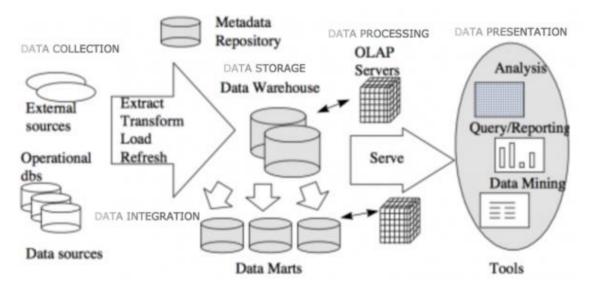
Figure 4 Business Intelligence Process (J. Wang et al., 2005) (Azeroual & Theel, 2018)

There are several case studies showing how a Business Intelligence system can benefit companies. In Hannula & Pirttimaki (2003) a study was carried out to find out what Business Intelligence represent for a Finnish large-scale companies in the year 2002. Telephone survey was used as a primary source to research methods the response rate reached 92%. One of the benefits found is that BI systems allowed a harmonization of the companies' way of thinking, broadening understanding of business in general. Also, it allows strengthening strategic planning, increasing professionalism in acquisition, and understanding the meaning of information among users. The results of this survey are shown in the figure 5. From this figure we can read the following benefits: better quality information acquired for decision-making (95%); improved ability to anticipate earlier the possible threats and opportunities (83%); growth of knowledge base (76%); increase of sharing information (73%); improved efficiency (65%); easier information acquisition and analysis (57%) and faster decision-making (52%).
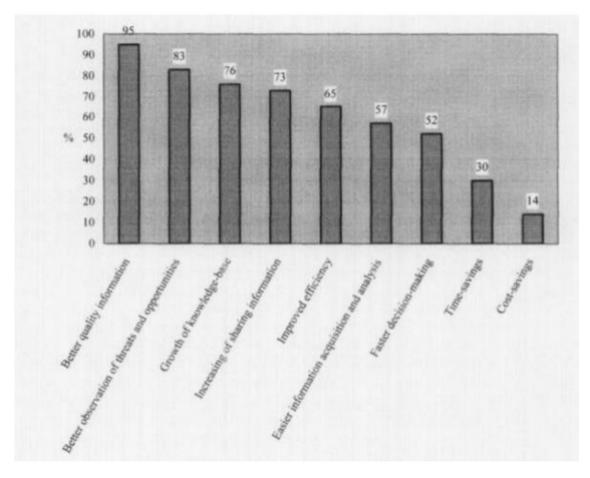
Figure 5 Benefits Expected from BI Activities (Hannula & Pirttimaki, 2003)

According to Barbara Wixom & Hugh Watson (2010) some Business Intelligence benefits are tangible and easy to measure. For example, companies may eliminate

software and hardware licenses and fees when they consolidate their tools, or companies may reduce headcount when they replace manual reporting processes. These kinds of benefits can be anticipated, tracked, and captured in a spreadsheet. Figure 6 illustrates the wide range of possible benefits resulting from Business Intelligence. The most tangible and easy-to-measure benefits have more of a local impact, typically happening at the departmental level. The more intangible benefits – things like process improvement and strategic enablement – can have impacts across the entire organization.
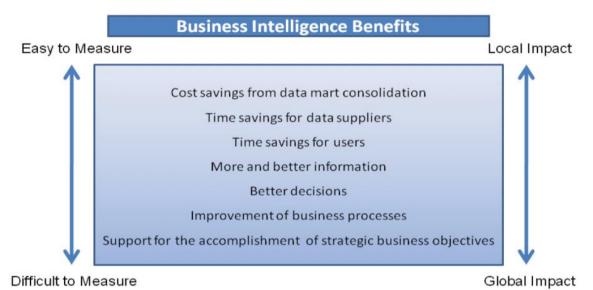
Figure 6 Benefits of Business Intelligence (Barbara Wixom & Hugh Watson, 2010)

On the other hand, Sentiment Analysis is "*the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes*" (Zhang et al., 2018) (Liu, 2020). In comparison to traditional surveys, Sentiment Analysis has some advantages. In the survey respondents are aware their answers will be analysed. While in Sentiment Analysis the respondents give their opinions without being aware their answers are going to analysed, being the interviewer a passive observer.

Although, being extremely advantageous, in Portugal only a small number of companies do Sentiment Analysis. The lack of development in this matter can be justified by several factors. The first is because web scrapping engines are not well developed in the Portuguese language. These systems are more advanced in the English language. Secondly, most Portuguese companies are small, as shown in Table 1: 99.9% are Small and Medium Companies. Small companies usually are lower profits and can invest less in social media (Organisation for Economic Co-operation and Development, 2000).

Table 1 Companies: total number and size (INE - Sistema de Contas Integradas das Empresas, 2021)

Year	Small/Medium	Big	Total
2017	1259234 (99,9 %)	1202	1260436
2018	1294037 (99,9 %)	1262	1295299
2019	1333649 (99,9 %)	1357	1335006

Thirdly, the lack of a technology-oriented culture. According to Dell Technologies (2018) report only 10% of Portuguese companies were consider Digital Leaders in 2018. Fourthly, marketing campaigns effects can be very difficult to measure. Unlike accounting for instance, marketing campaigns can ascribe feelings and emotions to customers (Bagozzi et al., 1999) (Matz et al., 2017). Lastly, not all Portuguese companies use social media and many make a deficient use of it (Guimarães, 2018).

So, is possible to say the incorporation of Business Intelligence and Sentiment Analysis can benefit companies. In the HEI case, Digital platforms allow institutions to not only publicize their work but also to get their feedback from the community. Keeping in touch with their alumni network and fostering conversations between the academic community.

Although, marketing is known to very difficult to measure, studying the impact of publications made in these platforms the allows the HEI to make batter decision on their communication plan and get to a wider public. As well as analysing the success of their communication strategy.

Having the contextualization in Business Intelligence and in Sentiment Analyses the next step is to study Digital Marketing and its relation to this work. The next section introduces social media and Digital Marketing.

2.2 Digital Marketing and Social Media

In this section the concepts of Digital Marketing and social media are described. Including a description of Social Media platforms used in this work, the tendencies in social media, the threats of Social Media Usage and social media in Higher Education Institutions.

Digital Marketing can be defined as a "set of techniques developed on the Internet with to persuade users to buy a product or service" (Avery et al., 2012). Although, this goal is

only achieved with the aid of a few tools such as Search Engine Optimization (SEO), Search Engine Marketing (SEM), as well as Social Media Marketing (SMM) (Saura, 2020). SMM happens when Digital Marketing is taken place in social media. On the other hand, social media can be described as "*a collection of software based digital technologies—usually presented as apps and websites-that provide users with digital environments in which they can send and receive digital content or information over some type of online social network*" (Appel et al., 2020). Another definition for this term is a "*Digital Marketing channel that marketers can use to communicate with consumers through advertising*" (Appel et al., 2020).

Digital Marketing platforms can have a serious impact in business. Studies have shown that social media can positively impact brand loyalty, making customers more loyal to the brand and increasing attention towards the brand (Erdoğmuş & Çiçek, 2012). Also, social media is proven to impact consumer's trust in the brand and a consumer who trusts the brand is more likely to buy products or services from it (Hajli, 2014).

In order to know how to take advantage of social media, companies must know its Social Media Background, their tendencies and threats.

2.2.1   Social Media background

In this section, Social Media Background is described for the platforms that are going to be data sources in the next chapter. Focusing on their foundation and business area. Also, some figures about the HEI profile in each social media are presented. These figures are related to the period of October 2021.

Social media has evolved throughout the years. The firsts well known worldwide social media platforms appeared in 2003, LinkedIn, MySpace, and Hi5. In 2004, Facebook, the social media with more users was created (Boyd & Ellison, 2007). In today's world social media is where users spend more time while online. Some studies refer users spend 4-6 hours per day on Social Media (Twenge et al., 2019) (Talaue et al., 2018).

Facebook is the world's most popular social networking, according to Statista, Facebook has roughly 2,89 billion monthly active users as of the second quarter of 2021 (Statista, 2021). It is noteworthy that Facebook is pointed as the preferred social media for educators (Chugh & Ruhi, 2018). In this platform users can post text-based messages, photos, or videos in their chronology. Users can like, share or comment other users' posts. Facebook and Instagram belong to the same group.

In the HEI where this research was conducted, there is a total of 21794 followers and a total of 21015 likes in this platform. The Facebook page was created in 2012.

Google Maps has a feature called "Local Guide" that enables users to write reviews and rate places and businesses and rate it from a scale of zero to five stars. The review can also have a comment which is text-based feedback of a business that customers leave. Other platforms have a similar business, like *Yelp, TripAdvisor* and *Dianping* (Mathayomchan & Taecharungroj, 2020). Is noteworthy that in this platform is not mandatory to write a text review and users can review the place they been only by rating it with the starts. The HEI had a total of 62 reviews in Google Maps and the average review score was 4.3.

Instagram provides users with the ability to share photos with others online. This platform is picture and video based (Hendrickse et al., 2017). Instagram has had a fast grown with up one billion monthly active users and more than 25 million business accounts (HubSpot & Mention, 2021). Research has shown that photos contain richer content in comparison with text based posts (Hu et al., 2014) (Salleh et al., 2016) (Ye et al., 2018) which can partly justly the brand success.

At Instagram, the HEI had 4791 followers, was following 5564 profiles and had 242 posts in this platform. The account was created in 2015, and in the same day the first post was posted.

LinkedIn is a social media dedicated to professional networking, job seeking, and recruitment, having over 660 million users in 200 countries (Forret, 2018). Studies shows that LinkedIn can be a good alternative to more traditional forms of face-to-face networking (Davis et al., 2020). The HEI had 12,649 followers in this platform. It was not possible to know when the page was created.

Forums are considered to be Online Communities that anyone can join to discuss topics of mutual interest (DeSanctis et al., 2003) (Biasutti, 2017). They allow "*participants to interact with text-based messages without time and place restrictions*" (Goggins & Xing, 2016). Also, "communications are posted, stored and preserved in a repository and are easy to access since they are ordered following specific discussion topics" (Q. Wang & Woo, 2008). One of the most relevant Student Forums in Portugal is Uniarea digital portal (composed of a website and a forum, as well as presence on social media) in which information on access to higher education in Portugal is disseminated (Gouveia, 2018).

At the local level is Forum Coimbra which is an online platform where students can post their questions. It is noteworthy that none of these platforms were managed by the HEI that is used in the present research.

Twitter is short text-based social media released in 2006 (History.com Editors, 2021). In this social media users can post tweets which are a "*message put on to let people know what the user is doing, thinking or feeling*" (Nuțu, 2019). According to Statista on the "*second quarter of 2021, Twitter had 206 million monetizable daily active users worldwide*" (Statista, 2021). This kind of posts can be retweeted which is to post something that has been written by one Twitter user and republished by another. It is noteworthy that Twitter is one the most popular social networks among students and young people (Shane-Simpson et al., 2018). The HEI had 552 followers on Twitter and was following 659 profiles. The year of the school's profile creation in Twitter was 2015.

YouTube was founded in February 2005 and, in 2007, the group *Alphabet Inc,* the same owner of Google, acquired YouTube (Arthurs et al., 2018) YouTube was the second most visited site in 2020 (J. Clement, 2021) and has been widely used for learning proposes. YouTube is the largest video sharing site on the Internet. In this Social Media users can post videos on channels and views videos from other channels. Also, users can comment, like, or dislike videos. The HEI's YouTube channel was created in 2011 and had a total of 69796 views. The number of subscribers of this HEI YouTube Channel is 386.

To conclude, figure 7 shows the years the HEI created a profile in each platform. The first platforms were YouTube and Facebook, in 2011 and 2012 respectively. In 2015 the profiles in Twitter and Instagram were created.
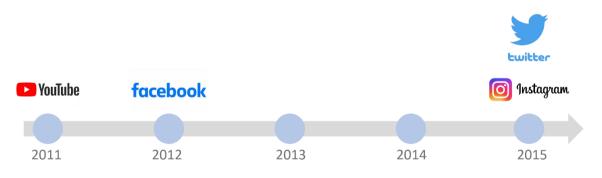
Figure 7 HEI Profile creation dates in Social Networks

Now that social medias platforms are presented in the next section some tendencies are described.

2.2.2 Tendencies in Social Media

In this section the latest tendencies in social media are described. For the business markets social media can be categorized as from of marketing. So, understanding tendencies is extremely important for companies who have an online presence. Allowing them to adjust content to each platform and get the most gain.

One Social Media tendency is the rise of social media platforms where images and videos replace text, for example Instagram and Snapchat (Appel et al., 2020). It is noteworthy that throughout the years users have passed over for image-based social media (Pittman & Reich, 2016).

The evolution in terms of domains of action can also be a tendency with the creation of more platforms specialized in one domain of action. For instance, LinkedIn for business and Trip Advisor for travelling. While others have worked on gaining other functionalities, like Facebook Marketplace, or WeChat's payment system (Appel et al., 2020).

Another tendency is the usage of celebrity endorsement strategy. This marketing strategy which is more common in consumer markets, consists on the usage of a well-known person who enjoys "*public recognition and who uses this recognition on behalf of a consumer product by appearing with it in an advertisement*" (McCracken, 1989). On the other hand, business markets usually use opinion leaders. The second ones, as being experts within an industry "*they carry information across the social boundaries between groups*" (Burt, 1999).

Social media tendencies vary with the time and companies must be aware of those changes to be relevant in social media. On the other hand, companies must be also aware of the threats social media can have which take us to next section.

2.2.3 The threats of Social Media usage

In this section the threats of Social Media usage are described. Despite, Social media is not the focus point of this work, getting to know its threats will allow a better understanding of this work's results and how these platforms are used. Social media allows users to connect and increase their visibility. Although, the usage of social media can bring some disadvantages, specially at the end-user level.

One threat is related to the end-user memory. Social networks usually promote a type of automatic attention, which causes instant pleasure without great effort from the individual. While learning, in its traditional sense, is recommended with the type of sustained attention, characterized by "*focus on performance on a single task over time*" (Esterman & Rothlein, 2019) in order to promote long-term memory (DeBettencourt et al., 2020). Long-term memory promotes knowledge and knowledge is intrinsic to critical thinking (Romiszowski, 2016) (Dwyer et al., 2014). So, by not saving the information on the long-term, the user is not able to master the topic widely and consequently s/he will never be possible to have a critical attitude about it.

Another concern around social media is privacy. It is known that social platforms have access to many users' data and this information is used for commercial proposes. For instance, it allows to personalize advertisements based the user's last web search (Pfiffelmann et al., 2020) (Jung, 2017). Some countries even created Data Protection laws. As an example, in 2016, the European Union adopted the General Data Protection Regulation (GDPR) which helped regulate the usage of users' data by companies and is based on the concept of privacy as a fundamental human right (Goddard, 2017).

As stated, many social platforms have invested in adding e-commerce functionalities or even creating their own payment system (Appel et al., 2020). With the COVID-19 pandemic the number of online purchases has increased (Pantelimon et al., 2020) as well as the number of cybercrimes (Ministerio da Administração Interna, 2019). Although not being only a problem of social media platforms, cybercrime also takes place in it. For instance, the Portuguese Annual Report of Internal Safety refers a growing the phishing, ransomware, and skimming crimes. On the other hand, some other crimes like sexual exploitation of minors are associated with social media, being the main source of its distribution (Ministerio da Administração Interna, 2019).

Another big threat of Social Media usage is the fast spread of fake news. Social Media are an important mean for the spread of fake news (Waszak et al., 2018). Some authors justified this with the presence of bots (Shao et al., 2018) (de Beer & Matthee, 2021). In the social media context, bots are software applications that "*can be programmed to search for information on the internet that is similar to what a social media user has already clicked on, liked, or shared*" (Burkhardt, 2017). According to Burkhardt (2017) users who get information via bots are more likely to share it even if haven not read them which may benefit fake news spread.

Fake news can affect brands significantly. For instance, brands like *The Pepsi Co, New Balance* and *Starbucks* have their image, sales and sometimes stock market value affected after misinformation campaigns (Obadă, 2019) (Domenico et al., 2021). To decrease fake news spread it is necessary an hybrid approach (de Beer & Matthee, 2021), with the use of digital literacy tools by the users and also Data Mining tools to identify these fake news (Figueira & Oliveira, 2017) (Shu et al., 2017).

Hate Speech is a "*speech that attacks a person or group on the basis of attributes such as race, religion, ethnic origin, national origin, sex, disability, sexual orientation, or gender identity*" (Nockleby, 2000) (MacAvaney et al., 2019) and is becoming more common in social media (ElSherief et al., 2018). The factors that contributed for this growing are the anonymity provided by these environments and increased willingness to express the users opinions online (Fortuna & Nunes, 2018). Text Mining plays an important role in preventing hate speech spread and help understanding its origin. Many articles talk about the way to detect Hate Speech. Hate Speech is often associated with Cyberbullying which leads us to next topic.

Since its appearance and widespread, social media has caused concern to parents, teachers, and mental health workers. This is due to several reasons related to teenage social media usage. One of them is cyberbullying, this is the type of bulling that takes place within online platforms "*such as emails, mobile, chat room, instant messaging, websites*" (Olweus & Limber, 2018) and happens repeatedly. In the recent years, an increase on the number of reported cyberbullying cases in Europe have been seen (Matos et al., 2018) (Livingstone et al., 2010).

Some authors even correlated social media usage with feelings of anxiety, depression, and suicidal thoughts. In Twenge & Martin, (2020) is shown that heavy users of digital media were often twice as likely when compared to low users to be low in well-being or have mental health issues, including risk factors for suicide. A more daily use is associated "*with a greater likelihood of participants scoring above the anxiety severity clinical cut-off indicating a probable anxiety disorder*" (Vannucci et al., 2017). So, it is shown that the more time spent using social media the greater are the symptoms of dispositional anxiety. The presence of high anxiety among social media users is also caused by the "*fear of missing out*" (FOMO). This is a social anxiety form that is characterized by the "*a fear that other people are having fun without you*" (Blackwell et al., 2017) (Przybylski et al., 2013) and is very often linked to the social media usage (Appel et al., 2020) (Hunt

et al., 2018) specially in younger adults. This phenomenon also helps Social Media Platforms to retain users regardless of their wellbeing.

Social network platforms are usually built by very large and profitable companies who are able to build extremely pleasant and easy to navigate websites and apps. These websites are extremely well thought out and configured to maximize the time of use without considering user's mental health. There are several strategies to ensure a longer usage time on these platforms (Neyman, 2017). However, these strategies have some disadvantages, especially for the end user.

To conclude, social media have its threats but there are ways to prevent it. Normally, the prevention is made by reducing the screen time. In Király et al., (2020), some recommendations are made like monitoring the screen time and placing the device somewhere not constantly available and turning off notifications. The threats of social media can be lessened by promoting information literacy amount users. Even children should be taught about the dangers of the internet and their activity must be monitored. Having concluded the contextualization about social in HEIs.

2.2.4 Social Media in Higher Education Institutions

In this section the usage of social media by HEIs is described, using some case studies. The usage of social media by HEIs presents a series of advantages to these institutions. Allowing HEIs to attract new students, to better communicate with the current students and staff, to keep in touch with former students, to foster conversations with other HEIs and to disseminate their work among the academic community.

Facebook is the social media with more users worldwide. So is important that HEI have a presence in this platform. In Peruta & Shields (2017) Facebook posts from 66 TOP United States colleges and universities were analysed. Data were extracted using Facebook's Graph API. Findings were that the lack of research in Digital Marketing focusing in HEI, make practitioners manage their Facebook pages based on "gut feelings" rather than substantiated best practices. Also, liberal Arts Schools had usually more engagement that other types of HEIs. Overall, HEIs post more photos than videos or text-based posts. Likes are the most important drive for total engagement. Schools posting more frequently see less overall engagement on each post.

Google maps reviews can be a good alternative to surveys, since the review is being made to the HEI and the user is aware his/her review will be read by others. Also, HEIs who

practice surveys are seeing low response rates (Shah et al., 2017) and with a small sample is difficult to make assumptions for the rest of the population.

Instagram is a recent and fast-growing Social Media platform, being very popular among young people. Although, it is proven that universities had a limited presence on Instagram and not fully use it to engage with current or prospective students (Stuart et al., 2017). In Almeida & Morais (2020) a descriptive study that allowed observing the Facebook and Instagram pages of 13 Portuguese Public Universities during September 2018. Findings were that posts focus mainly on welcoming new students and on the welcome messages delivered by the Presidents/Boards. Also, is shown that universities are increasingly using social networks to communicate with their students but still make a deficient use of social media. Their communication strategy could be improved by the use of Hashtags, by considering the relevant comments made by followers and by avoid sharing the same type of content over different platforms.

For HEI LinkedIn is a very important tool due to its nature focused on the professional networking and career development, therefore it is important for HEIs to be active users of this platform. Making the right usage of the platform can be challenging because of the lack of research focusing on the LinkedIn usage by HEI. Several studies shown that LinkedIn can be very useful to understand the career trajectory of former students (Lauder, 2013) (Daniels et al., 2021). Also, studies shows that LinkedIn users tend to be more introvert, with only a small fraction of the members actively participates by performing actions such as sharing, commenting or liking another member's content (Constantinov et al., 2019).

In Portugal the platform MOODLE is widely used by HEIs. This platform can function as an internal discussion forum, focusing in answering students' questions about the classes' content (Muñoz et al., 2017). Although, discussions forums external to the HEIs are more focused in deliberating about broader questions students have, especially when enrolling into a HEI. Studies have shown students who are thinking about enrolling higher education in Portugal are concern about the accommodation, with house prices increasing in the country, students struggle to find affordable accommodations (França et al., 2021).

In Veletsianosa et al., (2017) it was emphasized that academic institutions realize the importance of using Twitter for the dissemination of information, however they still make deficient use of this social network. For example, the hashtags used were mostly self-

references, not suggesting broader conversations. The importance of the institution's dimension in the number of followers on its Twitter profile was also evidenced, with these variables growing proportionally: as the number of students increases, also the number of Twitter followers. This study also highlighted the lack of control over the content of publications, which, with the objective of promoting institutions, may, sometimes, lead to an unrealistic and misaligned projection of the organization's objectives. It was found that academic institutions have, in general, difficulties when using Twitter to get involved in critical speeches, summarizing their action to dissemination and promotion content. However, the most relevant finding of this project was that, while the Tweets that promote the university receive some attention, the messages with the greatest attention are those that report a change in the university's normal day-to-day operations (Veletsianosa et al., 2017). Johnston et al., (2013) show that users prefer Twitter to be informed about events around the world, while using others networks like Facebook to keep in touch with friends and visit acquaintance profiles.

In Mccoy, Nelson, & Weigle (2017) some impact factors on the number of followers on academic institutions' Twitter accounts were examined. One of these factors is related to mentions on Twitter made by professors with high reputation, winners of Nobel prizes, for example. Another factor pointed out was the importance of sport in the number of followers, and institutions with a highly visible sports agenda tend to have a greater number of followers. Finally, the importance of having a Records Management plan was emphasized, to avoid duplication of accounts and keep track of changes.

Therefore in Linvill, McGee, & Hicks (2012) the different ways of establishing a communication on Twitter were analysed. Having a sample of 113 HEIs in the United States of America, their tweets were catalogued according to the principles of Kent and Taylor of dialogical communication (Kent & Taylor, 2002) which are creation of useful information, creation of visitors, conservation of visitors, dialogic loop. From this analysis it was found that the institutions are not using Twitter in a dialogical way and, instead, are using it mainly as an institutional news feed for the general public. Also noteworthy is the success of larger HEIs on this platform, which is due to the greater number of sporting events.

Studies have shown that YouTube can be used by HEI not only to publicize their work but also for academic learning. In Orús et al. (2016) is described the impact of the YouTube in a HEI class. This class was composed by first-year business management

and administration students from a major Spanish university. These students were invited to make audio-visual content in YouTube. Findings were that, after the video-making process, students had acquisition of cross-curricular competencies and a better final mark. Also, satisfaction levels among students increased.

To sum up, in this section was highlighted that, despite the lack of research in this scientific domain, HEI, in general, make a deficient use of social media, using it mainly to publicize their work and not to promote critical speeches. Also, it was brought into attention the importance of having different content in each specific platform. The importance of the HEI number of students and their sports teams are also importance factors in the HEI's number of followers online.

In the Table 2 there is a synthesis of the contributions stated in the research related with the use of social media in HEI which is categorized by social media, main objectives and main conclusions and recommendations/good practices.

Table 2 Synthesis of the research contributions

Social Media	Main Objectives and Main Conclusions	Recommendations/Good Practices
Facebook	-HEI use Facebook pages based on "gut feelings" rather than substantiated best practices, due to the lack of research in Digital Marketing (Peruta & Shields, 2017). -HEIs post more photos or text-based posts than videos (Peruta & Shields, 2017). -Schools who post more frequently see less overall engagement on each post (Peruta & Shields, 2017).	-
Google maps review	-Google maps reviews can be a good alternative to surveys. -HEIs who practice surveys are seeing low response rates (Shah et al., 2017).	-

Instagram	-HEIs have a limited presence on Instagram and not fully use it to engage with current or prospective students (Stuart et al., 2017). -Posts are mainly about the welcoming new students and on the welcome messages delivered by the Presidents/Boards (Almeida & Morais, 2020).	-Use Hashtags more (Almeida & Morais, 2020). -Consider comments made by followers (Almeida & Morais, 2020). -Avoid sharing the same content over different platforms (Almeida & Morais, 2020).
LinkedIn	-LinkedIn is useful to understand the career trajectory of former students (Lauder, 2013) (Daniels et al., 2021). -LinkedIn users tend to be more introvert and to not share, comment or like another member's content (Constantinov et al., 2019).	-
Student Forums	-Students who are enrolling in higher education in Portugal are concerned about the accommodation (França et al., 2021). -Forums are good platforms to answer students' questions (Muñoz et al., 2017).	-

Twitter	-Hashtags used by HEIs are mostly self-references (Veletsianosa et al., 2017). -The number of students has an impact on the number of Twitter followers a HEI has (Veletsianosa et al., 2017). -HEIs rarely get involved in critical speeches, summarizing their action to dissemination and promotion content (Veletsianosa et al., 2017). -Messages with the greatest attention are those that report a change in the university's normal day-to-day operations (Veletsianosa et al., 2017). -Users prefer Twitter to be informed about events around the world (Johnston et al., 2013). -Mentions by professors with high reputation and a highly visible sports agenda can lead to more followers (Mccoy et al., 2017). -HEIs use Twitter mainly as an institutional news feed for the general public (Linvill et al., 2012).	-HEIs should not use self-referential Hashtags (Veletsianosa et al., 2017). -HEIs should promote critical speeches (Veletsianosa et al., 2017). -HEIs should have a Records Management, to avoid duplication of accounts and keep track of changes (Mccoy et al., 2017).
YouTube	-Introducing a video-making process in classes can lead to a better acquisition of cross-curricular competencies satisfaction levels among students (Orús et al., 2016).	-

Having concluded the literature review is possible to start the first step of the work, the construction of the DW, which is in the next Chapter.

3   THE CONSTRUCTION OF THE DATA WAREHOUSE

In this chapter the steps to build the Data Warehouse, DW, are described. The steps are based on the ETL process. Starting with the Data Extraction tools followed by Multidimensional Model. Later takes place the extraction and transformation of data and then the data load. After the construction of the DW follows the results with Dashboards made with data from the DW.

A DW is described as a "*centrally managed and integrated database containing data from the operational sources in an organization*" (Moalla et al., 2017). The DW feeds the business intelligence system. There are several characteristics that an database must have to be considered a DW (W. H. Inmon, 2005). One of these characteristics is that DWs should be built around a single subject. In this case the DW design only has information extracted from social media in which the HEI has a profile. Also, they should collect data from multiple systems. This system collect data from several social media. Date should be considered. In this system almost all the extracted tables have a date field which allows to build a Date dimension in *Power BI*. Data should be non-volatile. This means that users cannot change the data and the DW should be able to be updated.

In the present research for the ETL process it was chosen *SQL Server* as a software to build this DW. One of the advantages of using this tool relies on the language it uses, which is SQL. This language is very commonly used in DWs. Also, according to the "Gartner Magic Quadrant for Operational DBMS"(Hardie, 2019) (Adrian et al., 2019) for 2019, Microsoft was the leading software in database management, as shown in the figure 8.

To conclude, having the DW built the next step is to make the Data Reports with this data which takes place in the point 4.
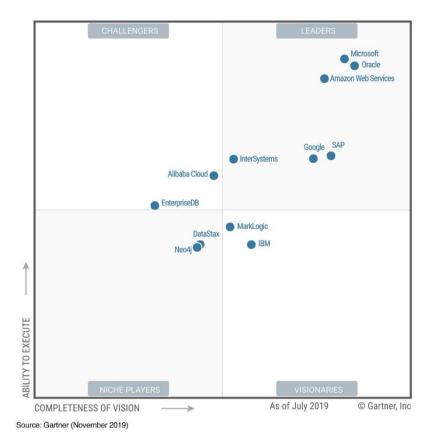
Figure 8 A Leader in 2019 Gartner Magic Quadrant for Operational DBMS (Hardie, 2019) (Adrian et al., 2019)

By constructing the DW in this tool the data can be accessed by another computer without the need of transferring files. Besides, queries can have an order of execution. Also, *SQL Server* has tool called *SQL Server Import and Export Wizard* that allows the integration of Excel and CVS files without the need of coding.

3.1   Data Extraction Tools

For the extraction of data, in this project, several software was used, changing most of the times according to the social media in analysis:

- For **YouTube** and **Twitter** an online platform called *botster (https://botster.io/)* was used. This platform allows to extract data based on the HEI's YouTube channel URL and the URL from the HEI profile on Twitter;

- For **Google Maps** the application used to extract data was *Instant Web Scrapper*. The extraction was based on the HEI's page in Google Maps;

- For the **Student Forums**, Uniarea and Forum Coimbra the *Power BI web connector* were used. This connector allows to extract data from online pages based on the URL. After the extraction the data can be exported to an Excel file.

- In **Facebook**, **Instagram** and **LinkedIn** data were directly extracted from the social media to several Excel files.

The table 3 has a summary of the tools used, *botster*, *Instant Web Scrapper*, *Power BI web connector* and Facebook, Instagram and LinkedIn to Extract, *SQL Server* to develop the tables and apply the transformation step and *Power BI* to build the Dashboards.

Table 3 Summary of the tools used

For the construction of the Data Report, *Power BI* were used. According to "Gartner Magic Quadrant for Analytics and Business Intelligence Platforms" (Dar, 2020) (Richardson et al., 2020) *Power BI* is the leading software for Business intelligence, as the figure 9 shows.

Is noteworthy that *Power BI* is a *Microsoft* tool for Business Intelligence, allowing the construction of Dashboards. *Power BI* uses DAX as programming language. For data transformations *Power BI* has a tool called *Power Query*.

Figure 9 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (Dar, 2020) (Richardson et al., 2020)
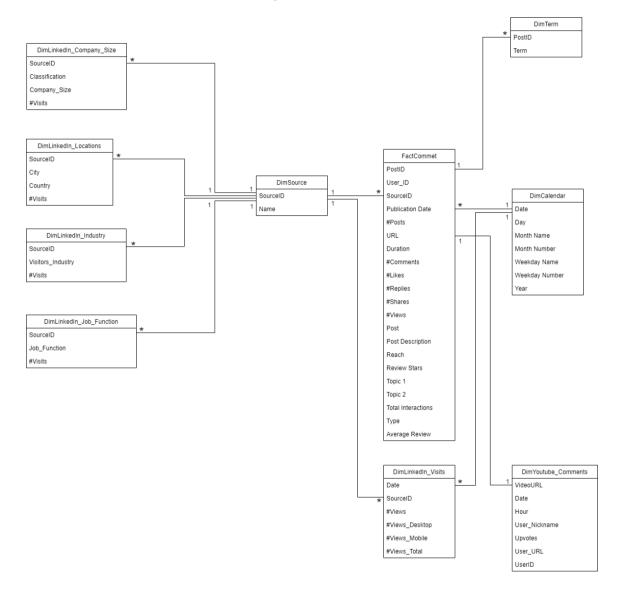
To summarize: several tools were used in this work, from the data extraction to the construction of the Dashboards. In the next section is described the multidimensional model that represents the information in the DW is described.

3.2 Multidimensional Model

In this section is described how the extracted tables were organized and their relationship. There are several ways to organize tables in the DW. The chosen way was through a Dimensional Model. This model has two types of tables, factual and dimensional (Golfarelli et al., 1998). Factual tables contain the numerical measurements of performance captured by the events associated with the business processes and are connected to the dimensional tables by key fields. These tables describe the events in analysis and usually occupy the biggest part of the model being strongly normalized and having many rows/records. Dimensional tables contain the descriptive attributes that give the context to measures: describing "who?", "what?", "when", "where?", "why?" occurred the events. Having less records, they occupy less space in the model being heavily denormalized.

In other words, factual tables contain the events while dimensional tables contain the characteristics of the events. The dimensional model has some advantages. One of them is that it provides data to business that are simple to navigate. Other advantage is that is allows an increase in query performance.

Tables can be organized by different schemas, being Star Schema, Constellation Schema and Snowflake Schema (Chaudhuri & Dayal, 1997). The most common and the one used in this work is the Star Schema. The figure 10 shows the final multidimensional model.

Figure 10 Multidimensional Model

In the centre is the factual table, FactComment, that is connected to DimSource, DimCalendar, DimTerm and DimYoutube_Comments. DimSource is connected to all LinkedIn tables. How this tables were built is explained in the next sections of this chapter.

3.3   Data Extraction and Transformation

In this section the Extraction and Transformation steps from the ETL process are described for each data source. In order to reduce the number of queries, in this work the extract and the transform steps were merged. Also, an example of the tables extracted is made. Is noteworthy that whenever a filed could compromise user's privacy the values were hidden.

As mentioned, the first step of the ETL process is the Extract. In this step comments, reviews, likes, and other information were extracted from several platforms in order to build the DW and load the information into the reporting software. It follows the transformation step, were transformations and cleaning are done to improve data quality.

The current work had some limitations in the extraction step. This work is based on single extractions and not directly connected to the data source, due to the academic nature of the work which was not funded to allow acquisition of licenses or other paid services provided by Social Media platforms. This also had an impact on the amount of information extracted from the platforms, which was lower.

Another limitation is the non-understanding of programming languages like Python, Node.js, PHP and C++ by the researcher. By not having a high knowledge of the previous languages no-code or low-code applications were chosen in the web scrapping phase, like *botster, Instant Web Scrapper*, *Power BI web connector.*

Also, the amount of information extracted from the different data sources is very heterogeneous since users are more active in certain forms of media. In the next sections the extraction and transformation from each data sources are described in more detail.

3.3.1   Facebook

The extraction from Facebook was the one with more limitations and problems of Data Quality. To extract data the first step is to login and choose the HEI Facebook page. Then is necessary to choose the option Page Insights and followed by Export Data. A pop-up will appear, and the users should choose a time window to extract an Excel file that contains data related to less than 180 days. So, the extraction was made into several Excel files because of this time restriction. Being the total of 17 individual extractions extracted with 1209 days. These files contain posts the HEI made.

These files had 7 tables with aggregated data. The table used in the report had information about the Post ID, a Permalink, Post Message with the description of the post, Type, weather it was Photo, Video, Link or Shared Video, posted with the date of publication of the post, shares with the number of total shares, like with the number of total likes and comment with the number of total comments. The table 4 shows an example of the tables extracted from Facebook, having the columns permalink and Post Message deleted to assure user's privacy. It is noteworthy that the fields Like, Shared and Comments are based on the post and all the shares the post had. For instance, if the post was shared once, the likes field is the total of likes is a sum from the post and the shared post.

Table 4 Example of the tables extracted from Facebook.

Post ID	Permalink	Post Message	Type	Posted	share	like	comment
231034673648307_4463202300431502			Photo	10/29/21 5:07 AM	1	3	
231034673648307_4463160250435707			Photo	10/29/21 4:52 AM		6	

All the individual extractions were joined in *Power BI* using the append function in *Power Query*. This function allows to join tables with the same field names. After the append tables were loaded into *SQL Server* via the tool *SQL Server Import and Export Wizard*. This tool allows to import and export data into and from *SQL Server* from a variety of formats, like Excel and CSV.

3.3.2   Google Maps Reviews

The great advantage of extracting data from this platform is that reviews are made about the HEI, while comments from social media are related to the post made.

The software used was *Instant Data Scraper* which is a fully functional browser add-on (Ciechanowski et al., 2020). This tool allows the extraction of the fields Profile name, post content, comments, likes that can be downloaded as an Excel file.

To extract the reviews the first step was to add *Instant Data* as extension and the tool's icon will appear in the browser. Then on Google Maps search of the HEI name and click on tool's icon, a table will be created, and the user is able to download it as CSV.

Table 5 shows an example of the table extracted from Google Maps Reviews. The table fields were the Review URL, the User URL, Username, User number of Reviews, User Picture, No starts, Review Date and the Review. Some of these fields, Review URL, User URL, Username, User Picture were hidden to assume users' privacy.

Table 5 Example of the table extracted from Google Maps Reviews

Review URL	User URL	Username	User number of Reviews	User Picture	No starts	Review Date	Review
			7		5	08/10/2021	Maravilhoso...Primeiro dia de aula.em MBA em Fiscalidade...FELIZ de mais!
			3		5	09/10/2018	Excelente escola com um corpo docente exemplar sobretudo na Area do Direito.

All the reviews at the time were extracted, being a total of 44 reviews.

3.3.3 Instagram

The extraction from Instagram takes places in Facebook. As mentioned, the two social media belong to the same group. After logging in Facebook and choosing the HEI page, to extract data is necessary to choose the option Page Insights, then See all insights and content. The user will be led to a page where s/he can choose the time window and extract the posts to an Excel file. Although, there were many constraints in the extraction, being only possible to extract 102 posts that according to Instagram had the most reach. Reach is the number of people who saw the post at least once (Facebook Business Help Centre, 2021). This extraction was made from July 2018 to August 2021.

The extracting resulted on a table containing information about, the post description, the post time, reach, the number of likes and the number of comments. The table 6 was an example of the table extracted from Instagram. The table fields were Post Description, Post time, Reach, Likes and reactions, Comments and Shares.

Table 6 Example of the table extracted from Instagram

Post Description	Post time	Reach	Likes and reactions	Comments	Shares
Aula Aberta | Honorários dos Agentes de Execução	2021-10-29T05:07:55	1102	3	0	1
Workshop | Preparação para a componente não letiva	2021-10-29T04:52:05	1790	6	0	0

3.3.4 LinkedIn

The data was extracted directly from LinkedIn. To extract data, users should login into LinkedIn, choose the option Analytics and then Export. A pop-up will appear asking the user to choose a time window, then data can be downloaded. Although, the maximum time window LinkedIn allows the extraction is the current year. The information extracted resulted on an Excel file with some sheets with aggregated information.

The first sheet listed the number of page visitors per city and country. Having a list of countries and then the number of visitors from these cities and countries. Also, a split by delimiter was done in Excel, so that the Location field would turn into two fields, city, and country. The delimiter was the comma and in Excel the function that allows the columns split is called Text to Columns. Later, this sheet was loaded into *SQL Server* resulting on a table named DimLinkedIn_Locations. The table 7 has an example of the sheet extracted from LinkedIn.

Table 6 Example of the table extracted from Instagram

Table 7 Example of the table extracted with locations extracted from LinkedIn

Location	Total views
Cologne Area, Germany	3
Madrid Area, Spain	5

The second sheet had information about the visitor's Job Function. Having a list of Job Functions and the number of visitors with those job functions. Some examples of job function are Business Development, Education and Operations. This sheet was loaded into *SQL Server* resulting in the table DimLinkedIn_Job_Function. The table 8 has an example of the table extracted from LinkedIn with visitor's Job Function data.

Table 8 Example of the table extracted with visitor's Job Function extracted from LinkedIn

Job function	Total views
Accounting	18
Administrative	2

The third sheet was related to the industry in which the visitors belong, having the number of visitors per industry. Some examples of industries are Accounting, Information Technology and Services, Marketing and Advertising. The data from this sheet resulted in the DW table called DimLinkedIn_Industry. The table 9 has an example of the table extracted from LinkedIn with visitor's industry data.

Table 9 Example of the table extracted with visitor's Industry extracted from LinkedIn

Industry	Total views
Computer Software	13
Telecommunications	2

The fourth table was the related to the company-size which the visitors belong, being the following groups some of the examples: 2 to 10 employees, 11 to 50 employees and 51 to 200 employees. The resulted table in the DW was called DimLinkedIn_Company_size. The table 10 has an example of the table extracted from LinkedIn with visitor's company-size data.

Table 10 Example of the table extracted with visitor's company-size extracted from LinkedIn

Company size	Total views
2-10	8
11-50	14

The last sheet had information about the number of views per day from January to July 2021. Having a list of days and the number of visits each day had. Two other columns were in this sheet. One with the number of visits made though a mobile phone and the second with the number of visits made with a computer. This sheet resulted in the DW table called DimLinkedIn_Visits. The table 11 has an example of the table extracted from LinkedIn with visitors by date data.

Table 11 Example of the table extracted with visitors by date extracted from LinkedIn

Date	Total unique visitors (desktop)	Total unique visitors (mobile)	Total unique visitors (total)
10/08/2021	9	15	24
10/08/2021	5	14	19

3.3.5 Student Forums

The Forums used were Uniarea and Forum Coimbra and both were extracted with the same mechanism, *Power BI web connector*. Being the extraction quite similar.

The first forum to extract information was Uniarea. The first step was to go the forum pager and using its search box write the name of the HEI. This resulted in 10 pages with

comments containing the HEI name. In *Power BI web connector*, the URL of the Uniarea page with the search, were inserted allowing the connection between *Power BI web connector* and Uniarea. With this was possible to have a table containing only posts from the first forum page. Although, the goal was to extract posts from all pages. So, more development was needed to extract all the posts and not only the first page. This second development was made in the *Power BI* tool, *Power Query*. The table containing information only for posts from the first page were loaded into *Power Query*. In the forum existed 10 pages with comments mentioning the HEI, each page had 20 posts. So, in *Power Query* a second table were built, this table had only a single column containing the number of the first post from each page, being 1, 21, 41, 61 ,81 ,101 ,121 ,141 ,161 ,181. This table was turned into a Custom Function in *Power Query* which allowed to extract all posts regardless of the page. The table 12 has an example of the table with data extracted from Uniarea.

Table 12 Example of the table extracted with data extracted from Uniarea

Date	Username	Topic 1	Topic 2	Post
26/07/2017		Candidatos 2017/2018	Quais as vossas 6 opções? (2017-2018)	Uma das minhas opções é solicitadoria, boa sorte! :D
08/08/2017		Candidatos 2017/2018	Previsões médias 2017/2018	Tenho média de entrada de 13,66 no curso de Solicitadoria e administração, e o último classificado do ano passado entrou com média 13. Acham que as médias podem subir muito este ano?

The process was repeated for the Forum Coimbra but considering this page had 3 pages with 15 posts per page, a total of 45 posts. The table 13 has an example of the table with data extracted from Forum Coimbra.

Table 13 Example of the table extracted with visitors by date extracted from Forum Coimbra

Date	Username	Post
15/09/2010		Calma é que está a ser dificil... se nao ficar na residença, nao tenho onde ficar..
08/08/2009		Apesar de eu não ser de Solicitadoria, penso que se fizeres um acompanhamento das aulas e da matéria te safas

The resulting tables was then exported to Excel to later be imported into the DW as shown in the point 3.4 of this work.

3.3.6 Twitter

For this social media, the platform *botster* were used. Giving the URL from the HEI's Twitter profile, information about the posts was extracted. The fields extracted were Username, User handle (a short of the username that allows mentions), Date of posting, Text, Retweet count and Like count as shown in table 14.

Table 14 Example of the table extracted with data extracted from Twitter

Username	User handle	Date of posting	Text	Retweet count	Like count
		Tue Jun 08 10:47:49 +0000 2021	Webinar | "As TIC ao serviço do Controlo de Gestão - Dashboards Integrados"	1	1
		Mon May 24 16:02:52 +0000 2021	Apresentação da Pós-graduação	0	3

This data was downloaded into an Excel file and then loaded on the Server. It is noteworthy that all posts made by the HEI in Twitter was extracted.

3.3.7 YouTube

For this social media the goal of the analysis is to understand what videos had the post impact. So, likes, dislikes, comments, and video information (title and description) were extracted from the HEIs YouTube Channel. The information was extracted for all videos posted in the social media.

Similarly, to Twitter the extraction of data from YouTube was made with the tool *botster.* The URL from the HEI YouTube channel was introduced in *botster.* Resulting in an Excel table with the fields: Channel URL, Video URL, Total views, Total likes, Total dislikes, Total comments, Video title, Video description, Publication date, Category, Video duration, Made for kids. The table 15 has an example of the table with data extracted from YouTube. Later this table was uploaded to the DW, resulting in the table FactYoutubeExtract.

Table 15 Example of the table extracted with data extracted from YouTube

Channel URL	Video URL	Total views	Total likes	Total dislikes	Total comments	Video title	Video description	Publication date	Category	Video duration	Made for kids
		342	8	0	0	Business Conference | Marketing e Sales Intelligence	Música neste vídeo: Jorge Palma - A Gente Vai Continuar - Antena 3	29 Dec 2020 15:23:49	Education	2M23S	false
		364	12	0	5	Presentation Movie		02 Dec 2020 16:23:56	Education	2H3M36S	false

Secondly the comments made by users in the HEI videos were extracted. To identify the videos with comments a SQL query was made. The code for the resulted table is in the Annex 1. In this query when this number of comments is smaller than 1 it means that the video does not have comments. This query delivers the URL for the 10 videos with comments. Based on this result other 10 extractions were made in *botster* to extract all the comments. These 10 extractions resulted in 10 tables that were joined through a UNION in *SQLServer*, since they all have the same fields (condition to perform a union). An example of the table extracted after the UNION command is in table 16. The resulted table in the DW is called YoutubeCommentsExtract has the following fields: Video URL, Nickname, Replies, text, Upvotes, User URL, Date.

Table 16 Example of the table extracted with data extracted from YouTube Comments

UserID	Video URL	Nickname	Replies	text	Upvotes	Date	Hour	User URL
			0	Love your stuff! Have a great day!	0	27/05/2019	5:59PM	
			0	Lição de história fantástica!	1	02/11/2019	12:09PM	

To conclude table 17 has a summarization of information extracted by social media, having the number of records, the dates from the first post and last post.

Table 17 Summarization of the extracted data.

Social Media	#records	First post Date	Last post Date
Facebook	5256	19/01/2016	20/08/2021
Google Reviews	44	01/10/2012	01/09/2021
Instagram	102	31/07/2018	08/08/2021
LinkedIn	-	01/01/2021	01/07/2021
Forum Coimbra	32	26/01/2008	15/09/2010
Forum Uniarea	159	12/08/2016	30/05/2021
Twitter	575	09/09/2015	08/06/2021
YouTube	100	25/03/2012	29/12/2020

The social media with more posts extracted was Facebook, with 5256 posts, followed by Twitter with 575 posts. The more recent posts were from Google maps Reviews, having the last post from 01/09/2021. In LinkedIn the extraction was based on the number of visits and not on the posts. While the oldest post is from Forum Coimbra, being from 26/01/2008.

3.4 Data Load

In this section, the load part of the ETL process is described. The model is formed by two types of tables, factual and dimensional as mentioned before. The factual table contains all the posts extracted and is directly connected to the dimensions DimUser, DimSource, DimCalendar and DimYoutubeComments. Also, it is connected to the FactTerm table.

The principal table is FactComment. This table contains information about the posts, the users, the source, the publication date. Having as fields PostID, SourceID, URL, No views, No Likes, No comments, Post, Post Description, Publication Date, Duration, No replies, Type, No Shares, Review Stars, Reach, Topic 1, Topic 2, Username and User ID. The table 18 shows the Fields from FactComment and an example of a value this fields could have.

Table 18 Fields from FactComment

Field	Exemple Value
PostID	231034673648307_2197162210368867
SourceID	1
URL	
No views	
No Likes	0
No comments	0
Post	Conferência	Eco-Escola
Post Description	
Publication Date	17/05/2019
Duration	
No replies	
Type	
No Shares	0
Review Stars	
Reach	
Topic 1	
Topic 2	
User Name	
User ID	

Being the factual table, it is in the centre of the model, having the other tables connected to it directly or via another table. This table was populated via the INSERT INTO statement which is used to insert new records in a table. A stored procedure, named

FactComent_Load, was created to populate the table FactComment. The code from the FactComent_Load stored procedure is available in the Annex 1

Another table is DimUser which is a dimension. This table has information about the user who made the post. Namely the fields UserID, Name, No reviews, User URL, User Image. Each row represents a user, and it is connected to the FactComment via the key field UserID. The extraction in certain data sources were made based on the HEI posts and not the posts made by users. In these cases, Facebook, Instagram, LinkedIn, Twitter and YouTube a key was created. A stored procedure, named DimUser_Load, was created to populate the table DimUser. The table 19 shows the Fields from DimUser and an example of values this fields could have. The code from the DimUser_Load stored procedure is available in the Annex 1.

Table 19 Example of values from DimUser

UserID	Name	No reviews	User URL	User Image
User - 142	User - 142			
User - 170	User - 170			

DimSource is also a dimension and contains information about the sources of the analyses. Unlike the other tables this table is not fed by other tables but manually through a stored procedure named DimSource_Load. The table 20 shows the DimSource. The code for the stored procedure is available in the Annex 1.

Table 20 DimSouce

SourceID	Name
1	Facebook
2	Instagram
3	Linkedin
4	Twitter
5	Youtube
6	Forum Uniarea

SourceID	Name
7	Forum Coimbra
8	Google

This table is also connected to another four LinkedIn tables. The first is DimLinkedIn_Locations that has information about the number of visitors per city and country. Allowing to know where the users who visited the institution's profile on LinkedIn are. The second table is DimLinkedIn_Industry. This table has information about the industry from the users who visited the HEI's profile on LinkedIn. The third table is DimLinkedIn_Job_Function having information about the work area of the users. The fourth table is DimLinkedin_Company_Size. This table has information about company size which visitors belong.

DimCalendar is a table made in *Power BI* using the DAX command CALENDARAUTO. This command returns "a table with a single column named "Date" that contains a contiguous set of dates. The range of dates is calculated automatically based on data in the model" (Microsoft, 2021). The code to build the DimCalendar and all the fields in the table are available in the Annex 2. Based on this column others calculated fields were created. The first one was the Month Number it to allows to sort the Month Name field based on the month order and not alphabetically. The second field created in this table was the Month Name and was created with the command swift based on the Month Number. Other fields were created, namely the Day, the Year, the Weekday number, Weekday name and Month-Year. Table 18 shows an example of the DimCalendar.

Table 21 Example of DimCalendar

Date	Month number	Month Name	Day	Year	Weekday number	Weekday name	Month Year
01/01/2008	1	Jan	1	2008	3	Tuesday	Jan-08
02/01/2008	1	Jan	2	2008	4	Wednesday	Jan-08

The calendar table is directly connected to tree other tables based on the date column. DimLinkedinViews has information about the daily number of views in the HEI LinkedIn profile. These views are spitted into mobile views and desktop views. Mobile views are

views made with a mobile gadget like a smartphone like LinkedIn table DimLinkedinViews. While Desktop views are made via a computer.

Another table is DimYoutubeComments. This table has information not about the posts made by the HEI in Youtube but about the comment's users made in those post. Being connected to FacComment by the field URL. Some of the fields of this table are, the users name, the comment, the number of upvotes and the date.

Lastly with all the terms presented in the extracted post another table was made. So, DimTerm represents each term that is in the posts. Having this tables allow us to understand what terms are more common. Also, this table is connected to the FacComment, that way it is possible to connect the terms to the comments and the other dimensions. To bult the table the field Post from FactComments were slipped, having as delimiter the space and the resulted columns were unpivot. So, each row represents a term used in the comments. Also, the table been through a process of cleaning in other to not consider stop words and special characters. Table 22 shows an example of two values from the DimTerm. The list of stop words were extracted form (Linguateca, 2006) and (GitHub, 2012) the code for this table can be seen in Annex 1.

Table 22 Example of DimTerm

PostID	term
231034673648307_1652702228148204	DIÁRIO
231034673648307_1876541865764238	COIMBRA

Once concluded the ETL process of constructing the DW the next step is to make the Dashboards with the information form the DW. So, next chapter is about the Dashboards built and their reading.

5   RESULTS – DATA REPORT AND DASHBOARDS

In this chapter the results are going to be presented and discussed according to each Social Media source, focusing on the reading and interpretation of the Dashboards made. Describing the topics that had most impact, the weekdays with more likes and the months with more likes or posts. Also, a Dashboard with the terms used at least twice in each platform was developed for all the data sources expect LinkedIn where it was not possible to extract posts.

Posts could be done by users, in the case of the Student Forums (Forum Coimbra and Uniarea), Google Maps Reviews and YouTube Comments. Or by the HEI, namely in Facebook, Instagram, Twitter and YouTube. Since the post publisher varies from users to the HEI, two dashboards were made. The first is in figure 11 and shows the terms for posts the HEI has made, excluding terms from the posts made by users. So, in this figure data is related to Facebook, Instagram, Twitter and YouTube. While in figure 12 is shown the terms used by users, excluding the terms used by the HEI.

In the figure 11 the terms used by the HEI in 4 platform are described. The most frequent used terms were Coimbra (1530), Diary (1545), Beiras (1113), the HEI name (941), Business (705) and School (624). It is noteworthy that Facebook had most posts. From this overall analysis is possible to understand the HEI uses these platforms to publicize its work, with terms related to its academic offer, like Accounting, Marketing, Courses, post-Graduation, Management and to inform users, with terms like Information, Info and inscriptions.

Figure 11 Dashboard "Terms used by the HEI"

On the other hand, figure 12 shows the terms used by users in Forum Coimbra, Uniarea, Google Maps Reviews and YouTube Comments. Users used more terms related to enrolling into Higher Education, with terms like Enrolling, Classes, Average score, Option and the School's academic offer referring to Management, Accountant, Administration, General Solicitor and Auditory.
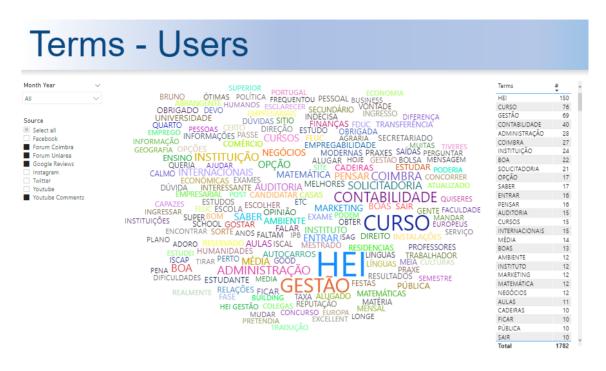
Figure 12 Dashboard "Terms used by users"

The next section describes in detail the results for each social media. User's name is anonymized and the HEI's name is covered.

5.1 Facebook

Figure 13 shows data extracted from Facebook. Having the posts made by the HEI by weekday, date and the list of all posts ranked based on the maximum number of likes. The first post is from January 2017 and the last from August 2021. Is noteworthy that the number of likes and the number of comments is an aggregated value. For instance, if a user shares the HEI's post the number of likes shown in the dashboard is a sum of the likes the post had in HEI's posts and in the shared post.

Figure 13 Dashboard "Facebook Posts"

In the pie chart is possible to understand which weekday had the most likes. The modal was Monday (21,2%), followed by Thursday (17.6%) and Friday (17.1%).

In terms of number of likes per date, is possible to say that this number is rising. In 2016 the total number of likes was 8524, while in 2020 the number was 45452. The same happened for the comments (290 in 2016 and 2657 in 2020) and for the number of shares (1366 in 2016 and 3688 in 2020). The month with more likes is May 2020 (7498 likes), mainly because of two posts. The month with more shares were February 2019 with 920 shares. In this month the post with more shares (366 shares) is about the launched of a new MBA. While the month with more comments is January 2021 with 567 comments. This number is justified by a post reporting the death of a teacher (430 comments).

The Facebook top 10 posts according to the number of likes had the following subjects: the school promotional video (2157 likes), two posts reporting the death of someone from school (grief) (1858 likes and 1619 likes), the school mascot (1529 likes), grief (1471 likes), reference to the marriage of two former students (1349 likes), presentation of the master's academic offer (930 likes), grief (834 likes), referring to the school mascot (761 likes) and tribute to health professionals during the pandemic (705 likes).

The second Dashboard is in the figure 14 and shows the most frequent terms used in Facebook posts by the HEI. The most 5 common terms are Diary (1545), Coimbra (1520), Beiras (1113), the HEI name (836), Business (651) and School (591). Is noteworthy that

the terms Diary and Beiras refers to the school's newspaper. So, is possible to conclude that posts made in this platform have a very small number of terms related to events outside of the school and posts are self-centred.

Figure 14 Dashboard "Facebook Posts-Terms"

To summarize, in this platform the subject is the promotion of the HEI's work, referring to school's projects and events. Posts that generate more likes were about reporting the death of a teacher or student, a reference to former students, about the school mascot, the school's academic offer and the pandemic. Also, the post with more likes was a video and Monday, Thursday and Friday were the days with more likes.

5.2 Google Maps Reviews

The figure 15 shows all the reviews made in the HEI page in Google Maps. Having a line chart with the average review and number of reviews by date, a pie chart with the number of reviews by weekday and a table with the review's date, the user's name, the URL to the user's profile, the review, the number of starts and the number of reviews user have made. The big advantage of his analysis is the fact that users are evaluating the HEI directly so there is no need to make assumptions based on posts.

The HEI only had 44 reviews in Google Maps. From these, only 11 had a review comment. So, from the 44 reviews, 33 users only rated the school and did not write a review. While 11 rate the school and wrote a review.

Figure 15 Dashboard "Reviews in Google Maps – Terms"

In the line chart is possible to analyse the number of reviews overtime. The number of reviews is growing, having 1 in 2012 and 22 in 2021. The month with more reviews was October 2017 with 10 reviews, followed by May 2021 with 5 reviews. About the monthly average number of stars, the minimum has 4 and the maximum was 5. Although, there was 6 reviews with less than 3 stars, 2 reviews with 2 starts and 4 reviews with one star. All the reviews with less than 3 stars did not have a comment. The weekday with more reviews were Sunday (29,5%), followed by Monday (27,3%). From the 11 reviews with a text comment, 8 praised the school's facilities, others praised the teaching and only one referred to an event promoted by the school. Also, only one post was related to the academic offer and work life.

From the figure 16 is possible to understand the most common terms users use to review the HEI. The more frequent terms are Coimbra (8), School (4), Building, Business, Excellent and Good (all shown 3 times). So is possible to say that most terms are related to a positive review and the school's building is an aspect to highlight.

Figure 16 Dashboard "Reviews in Google Maps – Terms"

To sum up, overall school's reviews were very positive, with the average score no less than 4 in each month. Users were more likely to post reviews on Sunday and Monday. The school's facilities and the teaching are strengths. Since, any of users who reviewed the HEI with less than 3 stars did not wrote a text review is not possible to study the aspects to be improved.

5.3 Instagram

Since Instagram has many constraints in data extraction, the total posts extracted from this platform were "the top 102 with a higher reach" as shown in Figure 17. Having a pie chart with the likes by weekday, a line chart with the number of likes and comments by publication date and a table with the publication date, the post, number of likes and number of comments.

Figure 17 Dashboard "Instagram top Posts"

About the weekday analysis, in the pie chart, the modal value was Saturday, with 3875 likes which represents 22% of the total. The modal value is similar to other social media, with was on the weekend. Although, in Instagram the second weekday with more likes was a workday, Monday with 3159 likes with represents 18% of the total. This is mainly due to a post with 751 likes. The third day with more likes was Friday with 3038 likes which represents about 17% of the total.

For the evolution analyses, from the line chart, is possible to say that the profile is growing in number of likes, having a total of 2425 likes in 2018 and 7435 in 2020. The month with more likes was October 2020 with 1416 likes. Overall, January and February were the months with less likes. The months with more comments were August 2020 and March 2021 with 15 comments. Followed by September 2019 and May 2020 with 12 comments in both months.

From the post analysis, in the table, is possible to understand the posts that had more impact based on the likes. Understanding the subject of this post can be a good input for the Communication Office responsible for managing this social media. The post that had the most likes (905 likes) refers to the school's mascot. The second post refers to going back to school after a lockdown caused by the COVID Pandemic (751 likes). The third refers to welcoming new students (652 likes). The fourth, fifth and sixth posts with more

likes refers to grief, reporting the death of someone from school, with 605,576 and 567 likes respectively.

The figure 18 shows the terms used in the Instagram posts by the HEI. In this platform the most frequent terms were the HEI name (31), Coimbra (28), Business (20), School in English and Portuguese (17 and 15 respectively), Conversations and Loggia (both 13 times). The term Loggia refers to one of the headings of a local newspaper. By the terms is possible to understand the school uses this social media mainly to publicize events related to the school.
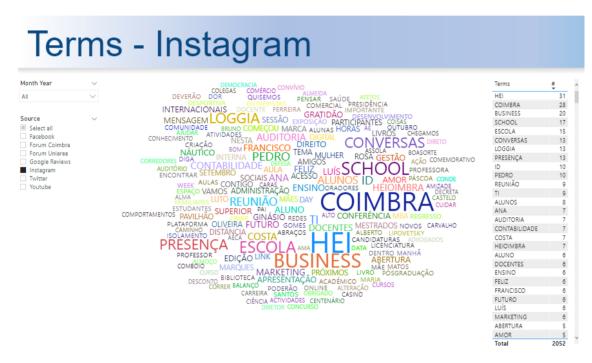

Figure 18 Dashboard "Instagram top Posts – Terms"

To conclude, the HEI uses this platform mainly to promote its work. Posts with more likes were about the school's mascot, the COVID Pandemic, welcoming new students and reporting the death of someone from school. Saturday and Monday were the days with more likes.

5.4   LinkedIn

Figures 19 shows data extracted from LinkedIn. Two Dashboard were made. The report is built according to unique visitors. Meaning that if a visitor made two visits only the first one counts. The first Dashboard is in figure 19 and shows information about users who visited the HEI LinkedIn page. Having 3 pie charts, visitors by job function, visitors by weekday and visitors by company size. The Dashboard also was a line chart with

visitors by date, making the split between mobile and desktop visitors and a bar chart with visitor's industry. Is noteworthy, that the extraction was made with 6 individual tables with aggregated data. So, it was not possible to connect the source tables.
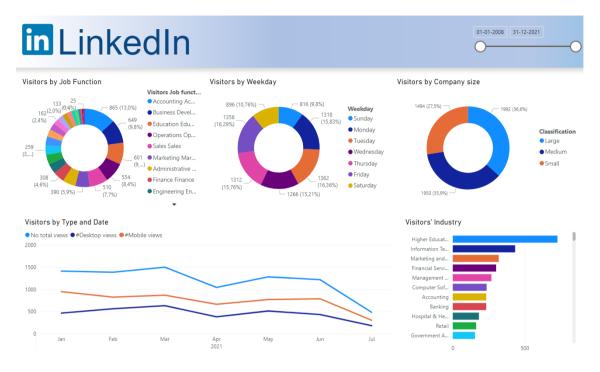

Figure 19 Dashboard "LinkedIn Visitors"

From the Job function pie chart is possible to understand the 5 most common visitor's job functions. The job function with more views were Accounting, 865 (13%), followed by Business Development (649 views (9,8%)), Education (601 views (9%)), Operations (554 views (8,4%)) and Sales (510 views (7,7%)).

From the days of week by visitors' analysis is possible to say that visitors were constant during the days. Except for the weekend (Saturday and Sunday) which had the less visitors. One reason for this pattern is that LinkedIn is dedicated to professional networking and the typical work week is from Monday to Friday. All the workdays had around 15% to 16% of the total visitors.

LinkedIn has 9 company size categories, varying from 1 to more than 10001 employees. The category with more visitors were the 51-200. The second one with more visitors were 11-50. The third one with more visitors were more than 10001. The fourth one with more visitors were 201-500. The fifth one was 10001-500. It is noteworthy that most Portuguese companies are medium or small companies (INE - Sistema de Contas Integradas das Empresas, 2021). With the provided classification data was very heterogeneous and did not have a pattern. So, a custom field was created to group

LinkedIn categories to other bigger categories. The created categories were inspired in the Commission Recommendation concerning the definition of micro, small and medium-sized enterprises (The Commission of the European Communities, 2003). So, three categories were created Small (less than 50 employees), Medium (less than 500 employees) and Large (less than 500 employees). Given by the DAX statement is available in the Annex 2. With these new categories it was possible to understand that users from larger companies were the ones that visited the most the HEI profile in LinkedIn, with 1992 (36,6%). Followed by Medium 1950 (35,9%) and Small 1494 (27%).

From the Visits by Type and Date, in the line chart, is possible to understand that visits from mobile are overall higher than visits from desktop. This difference is highlighted in the month with more views. In March 2021 the page received 1500 visitors, 868 from mobile and 632 from desktop. The second month with more visitors was January 2021 with 1412 visitors, 948 from mobile and 464 from desktop. Also, overall visits are decreasing with 1412 in January 2021 and 1220 in June 2021.

In the bar chart is also possible to analysis the Visitors' Industry. The industries with more views were: Higher Education (728 visitors), Information Technology and Services (434 visitors), Marketing and Advertising (320 visitors), Financial Services (302 visitors), Management Consulting (269 visitors). It is noteworthy the relation with the visitor's industry and the school's area which is Business Management.

Figure 20 has information about the visitor's location, being possible to understand the countries which visited the most HEI profile.

Figure 20 Dashboard "LinkedIn Visitor's locations"

Besides Portugal, with the most visits (572) the foreign country with more visits was Brazil with 589 visits. The following countries were Spain (77 visits), United Kingdom (48 visits), Italy (45 visits), France (32 visits), Germany (26 visits), Belgium (22 visits), India (22 visits), Switzerland (16 visits) and USA (16 visits). Overall, the HEI profile in LinkedIn was visited by users from 32 distinct countries.

To conclude, the number of visits is decreasing. Users tend to visit the HEI LinkedIn profile during workdays with little variance between the days. Visitors from large companies were the ones that visited the most. In terms of visitor's jobs and industry, users with functions related to business management visit more the HEI profile. From the location analysis is possible to understand the factor language (in the Brazilian case) and the factor to be in Europe were the most important.
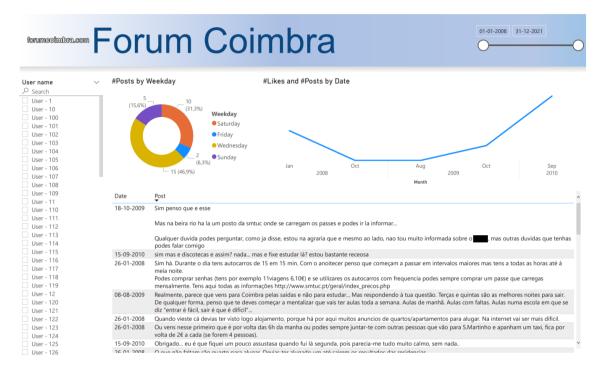
5.5   Student Forums

In this section the results from the student forums are described separately. Although, having the same extracting method the results were different among these two platforms.

Is noteworthy the information extracted from the Forums have a high value because posts are made by the students and shows their concerns regarding the HEI.

5.5.1 Coimbra Forum

Forum Coimbra had very dated posts, with the lasted post in 2010 and the first post in 2008. This may affect the analysis since the subjects change overtime. The total posts were 32 posts. Figure 21 shows the posts made in Forum Coimbra. Having pie chart with the number of posts by weekday, a line chart with the number of posts by date and a table with date and the post.

Figure 21 Dashboard "Forum Coimbra posts"

In the pie chart is possible to see the weekday that had, being Wednesday, Saturday, Sunday, and Friday. The two weekdays with more comments were Wednesday (46.9%) and Saturday (31.3%).

By analysing the posts in the table, is possible to understand that the most common topic was transportation, with 14 posts. Posts with this subject were for instance, questions on transportation schedule and which bus lines to take to go to school. Another topic was related to accommodation, users asked questions on accommodations prices and where were the areas with more transportation. One common topic was student life and gathering, with users asking questions about nightclubs in the city, for instance. Having 11 posts with this subject.

The figure 22 shows terms users used in Forum Coimbra. One of the subjects in this platform were accommodation with terms like Rent, Bedroom and Accommodation.

Other subject was transportation with terms like buses ("autocarro", "camioneta" in Portuguese) and bus passes ("passe" in Portuguese). Also, parties were a subject with terms like get out ("Sair") and parties ("Festas" in Portuguese).

Figure 22 Dashboard "Forum Coimbra posts – Terms"

To conclude: in this platform posts were very dated and in small number but shown some concerns students had, namely the accommodation, transports and social events. The days with more posts were Wednesday and Saturday.

5.5.2   Uniarea

Figure 23 shows the comments made in Uniarea. A bar chart with the number of posts by topic. A pie chart with the number of posts by weekday and a table with Date, the Topic 1 and Topic 2 and the Post. All the posts extracted from Uniarea mention the HEI and posts were not extracted according to their original order.

Figure 23 Dashboard "Uniarea comments"

Uniarea organize the conversations happening in the platform by topics. In the bar chart is possible to see posts by topics. The topics with more comments were the HEI (57 comments), the admission of new students and the list of candidates had together 29 comments, economics and business sciences (10 comments) and changing course (7 comments).

About the weekdays, seen in the pie chart, is possible to see that unlike the other platforms, the weekday with more likes was not a weekend day, being Tuesday with 18,2% of the comments. Followed by Thursday (17,6%), Wednesday (16,4%), Friday (13,8%), Sunday (13,2%), Monday (11,3%) and Saturday (9,4%).

In the table is possible to see posts made by users. Some of the subjects of these posts are questions regarding scholarship, the process to change courses and enrolling into higher education.

In figure 24 is show the terms used in Uniarea by users. The terms more used were the HEI name (141), Course (75), Management (67), Accounting (39) and Administration (28). Is noteworthy that the last 3 terms are also the names of HEI's courses. Users use this platform mostly to get information about enrolling in higher education with terms like Course, Option, Average, enrolling were being used.
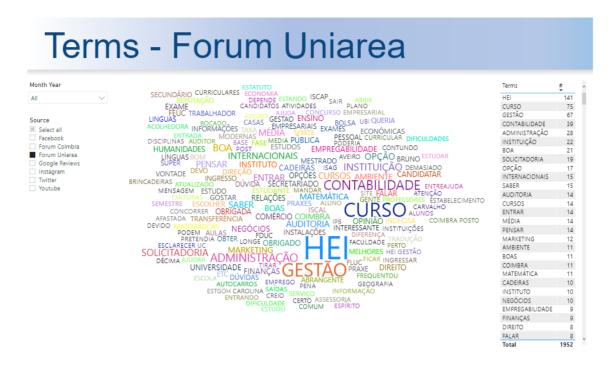
Figure 24 Dashboard "Uniarea comments – Terms"

To summarize, is possible to say this platform is mainly used by users to obtain information they need by asking questions. The main subjects are about the process to change courses, enrolling into higher, scholarship, economics and business sciences. Working days like Thursday, Wednesday and Friday were the days with more posts.

5.6 Twitter

Figure 25 shows a line chart with the number of likes and shares from tweets by date, a pie chart with likes by Weekday and a table with the publication date, the tweets and the number of likes and shares. The table is ranked, starting with the tweet with most likes.

Figure 25 Dashboard "Twitter Posts"

From the line chart is seen that the number of likes has increased substantially in 2020. Having the modal value in March 2021 with 140 likes and 58 shares. Followed by October 2020 with 132 likes and 44 shares. This is justified by the posts made which leads us to the next visualization, a table with all the posts, date of publish, likes and retweets. To understand which subjects had the most impact in the HEI's profile on Twitter the top 10 posts with more likes should be explained. The first post with 124 likes were about grief, reporting a dead of a student.

The second post with more likes (87) is regarding the coming back to school after a period of lockdown due to the Covid Pandemic. The third post with more likes (45) were about a conference with the school's presentation. The fourth post with more likes (36), were an image with a message about the HEI's birthday. The fifth post (36 likes) were about the opening of a new curse. The sixth and seventh (with 16 and 11 likes respectively) were about Christmas and new year. The eighth (10 likes) were about welcoming new students. The nineth (10 likes) was a message of hope about the Pandemic. The tenth (10 likes) was about the first women to graduate from a university. It is noteworthy, that for this top when the number of likes were the same, the number of shared counted to break the tie.

The other visual is pie chart with the number of posts per day of the week. Overall, the weekdays with more likes were Friday (24.1%), followed by Tuesday (20.99%).

In figure 26 is possible to see the terms used in the HEI's posts in Twitter. The most frequent terms were Information (170), Course (58), Management (47), post-Graduation (32), Free (30) and Business (29). Is noteworthy that Portuguese HEI have a type of academic offer called "Free Training" which can justify the term "free" being one of the most common. By the used terms is possible to understand the school uses this platform mainly to publicize its academic offer.
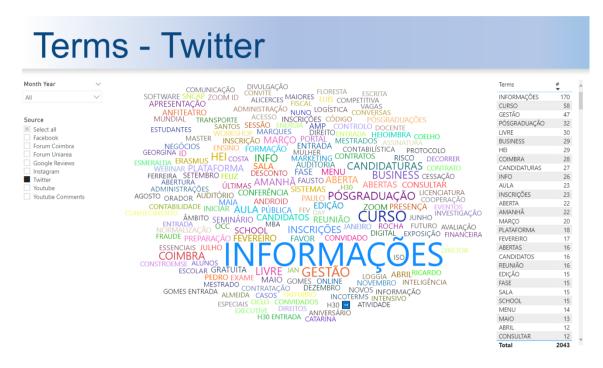
Figure 26 Dashboard "Twitter Posts – Terms"

To conclude, on twitter the top ten posts with more likes were very heterogeneous in subjects and number of likes. The most successful posts had as topics: grief, the pandemic and welcoming new students. From the terms analysis is possible to understand the HEI uses this platform mainly to publicize its work and to inform about the academic offer. Also, weekend posts were the ones with more likes.

5.7 YouTube

Figure 27 shows information about videos posted in the HEI YouTube channel. Having a line chart with the number of Views, Likes and Comments by date and two tables. One with information about the video, Publication Date, URL, Video, Video Description, Number of Views, Likes and Comments. The second one with the comments made in the videos, Date, URL, and the Comment.

Figure 27 Dashboard "YouTube Posts"

Unlike the other platforms where the likes are the most important indicator, in this social media, the most important indicator is the number of views.

In the line chart is possible to understand that summer months were the ones with more views. The month with more views was July 2012 with 17503 views, 46 likes and 2 comments, followed by September 2017 with 10571 views, 240 likes and 12 comments. Also, the difference between months is very high, with June 2012, for instance, with 661 views. From the analysis of this chart is possible to say the HEI is not investing in this platform. The last post was from December 2020. In 2019 and 2018 only two posts were made.

In the table is possible to see the videos with more views. The post with more views (17503) was a promotional video with the school presentation. The second video with more views (7425) was a presentation with an opinion leader and former politician. The third post (6490) were a presentation also with an opinion leader and law professor. The fourth (3792) post were a school promotional video but from another year. The fifth (3533) post was a presentation also with an opinion leader and teacher. The sixth (2839) video was a debate about Democracy and Corruption. The seventh (2367) post was a school promotional video. The eighth and nineth videos were also with opinion leaders with 1657 and 1371 views respectively. The tenth post (880) was a promotional video.

In figure 28 the terms used in the HEI's video titles are shown. The more frequent terms were the HEI name (45), Minute (34), Utopias (24), Real (10) and Business (5). Is noteworthy Minute and Utopias are two headings the school had in this platform.

Figure 28 Dashboard "YouTube Posts – Terms"

Since YouTube videos comments were extracted a second Dashboard with terms used by users was made and can be seen in the figure 29. In this social media terms related to politics were used, like Europe, Portugal, Politicians, Rights, Ethics, Freedom and History. So, in this platforms user spoke more about politics. It is due to the school's promoting debates with this subject.

Figure 29 Dashboard "YouTube Comments"

To sum up, in the school promotional video and presentations with opinion leaders were the posts with more views. The school is not investing in this platform. Many posts only had the video title and no description. Unlike other social media profiles from the HEI, this platform was used mainly to promote the discussion of ideas with opinion leaders.

6   RECOMMENDATIONS

In this section recommendations to the HEI Communication Office, based on the results presented in previous chapter are made. Overall, we may state that the strategy used by the HEI is effective. Facebook has a feature named Pages to Watch that allows the comparison to other similar pages. This comparison is made based on total Page likes, the number of posts published and the amount of engagement (likes, comments, and shares) (Cherry, 2018). The pages Facebook compare the HEI, belong to schools with a higher number of students and from highly populated cities. Also, a comparison in Facebook to two other schools with a similar academic offer, organizational structure and within 100 km from the HEI was made. The first competitor had a Facebook profile since 2010 and had 13872 page likes. The second competitor had a Facebook page since 2011 and 3228 page likes. So, in comparison with its competitors the HEI more followers in social media

Although, there are a few recommendations. HEI should post more in times that with more traffic, like the weekend. Post about subjects that have are known to generate many impressions, like the school mascot, the pandemic and welcoming new students. Also, the HEI should maintain a record of the posts, in order to avoid duplication and keep track of changes (Mccoy et al., 2017).

Facebook, Instagram and Twitter posts that the HEI makes are mainly to publicize its work. Is recommended to have broader and current subjects in future posts, promoting discussion among users. For instance, posts referring to recent well-known events or social ideals. The HEI has already done at least one post with this characteristic, referring to Gender Equality when enrolling into Higher Education.

This work allowed to understand, in detail, users who visited the HEI's LinkedIn profile. Visitors from large companies and who work in business management are the ones who visit more the HEI profile. Also, Portuguese speaking and Europeans users are more likely to visit the profile. So, posts aiming these users could benefit the HEI, since they are the ones more interested. Posts promoting cultural diversity and students exchange and highly recommended.

From the Student Forums analysis, Coimbra Forum and Uniarea, it is possible to understand that students are concerned about accommodation and transports, since this HEI is not at the centre of Coimbra and less accommodation and transports options are available, specially concerning the Academic Events where undergraduate students

usually intend to join to. So, posts with these subjects would have a high value of information, contributing to lessen students concerns. Also, to have a constant monitoring in order to understand if students' concerns vary to problems other than accommodation and transport. This students' concerns should be shared with other school's departments.

In YouTube users seem to engage with the content created. Although is seen a disinvestment in this platform, with very few posts in the last years. Content in video form is known to provide a good user experience and information delivery. So, this platform is a good channel to get to users and more investment should be made by creating more content.

The table 4 has a summarization of the recommendations and the already good practices in the school's communication strategy.

Table 23 Recommendations and Good Practices in the HEI's communication strategy

Social Media	Recommendations	Good Practices
Facebook	-HEI should post more in times that with more traffic, like the weekend.	
Instagram	-Post more about school mascot, the pandemic and welcoming new students.	
Twitter	-Promote discussion among users, posting about well-known events or social ideals. -Post about accommodation and transports since. -Post more about ex-students.	-Post about School changes. -Short Text Posts. -Different posts according to each platform.
Google Maps Reviews	-Understand the reasons why the school's facilities and the teaching are strengths. Share this information with other school's departments.	

LinkedIn	-Post aiming Portuguese speaking and European's students.	
Student Forums	-Keep a constant monitoring of the forums, to understand if students' concerns vary to problems other than accommodation and transport. Share this information with other school's departments.	
YouTube	-Invest in YouTube. Creating more content in this platform and sharing this content in the other school's platforms.	

To conclude, the HEI' communication strategy is effective, and the number of likes is growing in most platforms. The HEI is compared with much bigger institutions by the platforms.

7 CONCLUSION AND FUTURE WORK

In this work a Business Intelligence System for social media in HEI was developed. Findings were that Twitter, YouTube, and Google Maps were the easiest to extract information and Facebook and Instagram were the most difficult Social Medias to extract information. While Twitter, YouTube, and Google Maps were the easiest to extract information. Posts that had the most likes were about: grief (reporting a death of someone from the school), the school mascot, the pandemic or welcoming new students. Overall, the weekends were the days with more interactions, but this varies according to the social media and the month. Google Maps Reviews, Instagram and Twitter had more interactions on the weekend. In the Students forums, Coimbra Forum and Uniarea, users choose subjects related to accommodation, transport, and the school academic offer. In general, the number of interactions is growing, especially in Facebook.

The HEI has a good communication strategy. Having different content according to each social media and by the right usage of Hashtags and not sharing the same content over different platforms (Almeida & Morais, 2020). The HEI competitors according to Facebook are schools with much bigger dimension and from big cities. Although, post made by the HEI are mainly to publicize the school's work and inform users, instead of promoting the discussion on general topics. Also, the process part which took more time was the data extraction.

7.1 Main contributions

Social media are a long-term tool for HEIs to publicize their work. Although, a very small number of companies are using sentiment analysis in Portugal, this work provides research in an area where there is little scientific production. This issue is not only related to HEIs but companies from all types. To sum up, this work allowed to prove that sentiment analysis can be done to any type of company regardless of its size and financial resources.

In this work was shown what kind of posts had the most likes, the more popular days and some student concerns. Recommendations were made to the HEI Communication Office in order to improve the communication strategy.

This work took place after two articles submitted and accepted in conferences. Being the first page of these articles in Annexes 3 and 4, respectively:

- B. Varela, J. Bernardino and I. Pedrosa, "Twitter Sensitivity Analysis in a Higher School Using Power BI," *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, 2020, pp. 1-6, doi: 10.23919/CISTI49556.2020.9140979.

- Varela B., Bernardino J., Pedrosa I. (2021) The Application of Text Mining for the Analysis of Connotation in a Higher Education Institution. In: Rocha Á., Reis J.L., Peter M.K., Cayolla R., Loureiro S., Bogdanović Z. (eds) Marketing and Smart Technologies. Smart Innovation, Systems and Technologies, vol 205. Springer, Singapore. https://doi.org/10.1007/978-981-33-4183-8_10

7.2 Limitations

In this section is described the main limitations this work had and when possible how they were overcome.

The current work had some limitations in the extraction step. This work is based on single extractions and not directly connected to the data sources due to the academic nature of the work which was not funded to allow acquisition of licenses or other paid services provided by Social Media platforms. This also had an impact on the amount of information extracted from the platforms, which was lower. Another limitation is the non-understanding of programming languages like Python, Node.js, PHP and C++ by the researcher. By not having a deep knowledge of the previous languages is has chosen no-code or low-code applications in the web scrapping phase. Applications like the *botster*

Instant Web Scrapper and *PowerBI web connector*.

Another limitation was the lack of literature produced about the extraction of data from certain platforms. Platforms like LinkedIn had very little literature produced about how to extract data from it. Other platforms like Facebook have very little information explaining what each field from the extracted tables represented.

The heterogeneity of data was also a limitation with fields names non-normalized. Each data source had a different name for the same information. For instance, for the field showing when the post was done (Publication Date), in Facebook the field name was "Post Date", in Google Maps reviews "time of publish" and in Twitter, "date".

The access to non-aggregated data was another limitation. For instance, data extracted from Facebook about likes and comments were based on the post and all the times this post was shared. In LinkedIn was only possible to extract the unique visits by day, visitor's locations, job function, industry and company size.

Time was also a very predominant limitation. Business Intelligence is very iterative, and the Data Extract part was very time consuming which took time from the report writing.

To conclude this work had some limitations but they were softened by choosing user-friendly tools like *botster*, *Instant Web Scrapper*, *Power BI web connector*. To soften the lack of literature several attempts were done. The limitations about the access to non-aggregated data had impact on the information shown in the Dashboards.

7.3   Future Work

Business Intelligence is always iterative and there is always room for improvement. So, there are many factors to consider as future work. One is a direct connection to the social media platforms. Making the extraction with automatic refreshes and in real time. Another one is to apply data analysis algorithms to the extracted data. Also, extracting data from other platforms related to education like *Research Gate*. Allowing to analyse the publications made by the students and college researchers. Some of the possible indicators is the number of publications per year, the author and the impact of these publications. Another Future work is to study only the users who follows the HEI profile in social media or have posted something related to the institution. By knowing its users, the HEI could personalize its message and deliver better value information. Also, to study the profile of former students can be a future work, understanding their career trajectory. Lastly, to study the relation between the HEI and bodies related to students, cultural events or newspapers. Analysing, for instance, when HEI's posts are shared by this bodies produce more reactions than those from the HEI itself.

REFERENCES

Adrian, M., Feinberg, D., & Cook, H. (2019). *Gartner Magic Quadrant for Operational Database Management Systems*. Gartner Research. https://www.gartner.com/en/documents/3975492

Almeida, E., & Morais, N. S. (2020). The Use of Facebook and Instagram in Communication with New Students: the Case of Portuguese Public Universities. *SIIE*.

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79–95. https://doi.org/10.1007/s11747-019-00695-1

Arthurs, J., Drakopoulou, S., & Gandini, A. (2018). *Researching YouTube*. SAGE Publications Sage UK: London, England.

Avery, J., Steenburgh, T. J., Deighton, J., & Caravella, M. (2012). Adding Bricks to Clicks: Predicting the Patterns of Cross-Channel Elasticities over Time. *Journal of Marketing*, *76*(3), 96–111. https://doi.org/10.1509/jm.09.0081

Azeroual, O., & Theel, H. (2018). The Effects of Using Business Intelligence Systems on an Excellence Management and Decision-Making Process by Start-Up Companies: A Case Study. *International Journal of Management Science and Business Administration*, *4*(3), 30–40. https://doi.org/10.18775/IJMSBA.1849-5664-5419.2014.43.1004

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, *27*(2), 184–206.

Barbara Wixom, & Hugh Watson. (2010). *The Bi-Based organization*.

Biasutti, M. (2017). A comparative analysis of forums and wikis as tools for online collaborative learning. *Computers & Education*, *111*, 158–171. https://doi.org/https://doi.org/10.1016/j.compedu.2017.04.006

Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences*, *116*, 69–72. https://doi.org/https://doi.org/10.1016/j.paid.2017.04.039

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Burkhardt, J. M. (2017). *Combating fake news in the digital age* (Vol. 53, Issue 8). American Library Association.

Burt, R. S. (1999). The Social Capital of Opinion Leaders. *The ANNALS of the American Academy of Political and Social Science*, *566*(1), 37–54. https://doi.org/10.1177/000271629956600104

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod Record*, *26*(1), 65–74.

Cherry, B. (2018). *Why "Pages to Watch" Is a Valuable Facebook Marketing Tool*. Bluleadz. https://www.bluleadz.com/blog/pages-to-watch-tool-for-facebook-marketing

Chugh, R., & Ruhi, U. (2018). Social media in higher education: A literature review of Facebook. *Education and Information Technologies*, *23*(2), 605–616.

Ciechanowski, L., Jemielniak, D., & Gloor, P. A. (2020). TUTORIAL: AI research without coding: The art of fighting without fighting: Data science for qualitative researchers. *Journal of Business Research*, *117*, 322–330.

Constantinov, C., Popescu, P., & Mocanu, M. (2019). Identifying LinkedIn Usage Patterns Based on Feedback from Students and Fresh Graduates. *2019 18th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, 1–6. https://doi.org/10.1109/ROEDUNET.2019.8909507

Daniels, R. A., Pemble, S. D., Allen, D., Lain, G., & Miller, L. A. (2021). LinkedIn Blunders: A Mixed Method Study of College Students' Profiles. *Community College Journal of Research and Practice*, 1–16. https://doi.org/10.1080/10668926.2021.1944932

Dar, P. (2020). *Gartner's 2020 Magic Quadrant is Out*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/02/gartners-2020-magic-quadrant-bi-analytics-tools/

Davis, J., Wolff, H.-G., Forret, M. L., & Sullivan, S. E. (2020). Networking via LinkedIn: An examination of usage and career benefits. *Journal of Vocational Behavior*, *118*, 103396. https://doi.org/https://doi.org/10.1016/j.jvb.2020.103396

de Beer, D., & Matthee, M. (2021). *Approaches to Identify Fake News: A Systematic Literature Review BT - Integrated Science in Digital Age 2020* (T. Antipova (ed.); pp. 13–22). Springer International Publishing.

DeBettencourt, M. T., Williams, S. D., Vogel, E. K., & Awh, E. (2020). Sustained attention and spatial attention distinctly influence long-term memory encoding. *BioRxiv*, 2020.09.14.297341. https://doi.org/10.1101/2020.09.14.297341

Dell Technologies. (2018). *Digital Transformation Index*. Dell Technologies. https://www.delltechnologies.com/en-us/perspectives/digital-transformation-index.htm#scroll=off&pdf-overlay=//www.delltechnologies.com/asset/en-us/solutions/business-solutions/industry-market/dell_technologies_digital_transformation_index_ii_full_findings_report.pdf

DeSanctis, G., Fayard, A.-L., Roach, M., & Jiang, L. (2003). Learning in Online Forums. *European Management Journal*, *21*(5), 565–577. https://doi.org/https://doi.org/10.1016/S0263-2373(03)00106-3

Direção-Geral de Ensino Superior. (2021a). *1ª fase do Concurso Nacional de Acesso ao*

ensino superior público 2021 - Número de candidatos ao ensino superior volta a crescer e atinge novo máximo de cerca de 64 mil. https://wwwcdn.dges.gov.pt/sites/default/files/nota_cna2021_1afase_v20agosto_20 21.pdf

Direção-Geral de Ensino Superior. (2021b). *Ensino Superior em números.* https://www.dges.gov.pt/pt/pagina/ensino-superior-em-numeros

Directorate General for Higher Education. (2021). *Portuguese Higher Education System.* https://www.dges.gov.pt/en/pagina/portuguese-higher-education-system

Domenico, G. Di, Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341. https://doi.org/https://doi.org/10.1016/j.jbusres.2020.11.037

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52. https://doi.org/https://doi.org/10.1016/j.tsc.2013.12.004

ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., & Belding, E. (2018). Hate lingo: A target-based linguistic analysis of hate speech in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1).

Erdoğmuş, İ. E., & Çiçek, M. (2012). The Impact of Social Media Marketing on Brand Loyalty. *Procedia - Social and Behavioral Sciences*, *58*, 1353–1360. https://doi.org/https://doi.org/10.1016/j.sbspro.2012.09.1119

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. https://doi.org/https://doi.org/10.1016/j.copsyc.2019.03.005

Facebook Business Help Centre. (2021). *Reach | Facebook Business Help Centre.* Facebook for Business. https://www.facebook.com/business/help/710746785663278

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, *121*, 817–825.

Forret, M. L. (2018). *Networking as a job search and career management behavior. InU. C. Klehe & EW van Hooft.*

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, *51*(4), 1–30.

França, T., Cairns, D., Calvo, D. M., & de Azevedo, L. (2021). Lisbon, the Portuguese Erasmus city? Mis-match between representation in urban policies and international student experiences. *Journal of Urban Affairs*, 1–15. https://doi.org/10.1080/07352166.2021.1950544

GitHub. (2012). *Portuguese stop words .* https://gist.github.com/alopes/5358189

Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European

Regulation that has a Global Impact. *International Journal of Market Research*, *59*(6), 703–705. https://doi.org/10.2501/IJMR-2017-050

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education*, *94*, 241–251.

Golfarelli, M., Maio, D., & Rizzi, S. (1998). Conceptual design of data warehouses from E/R schemes. *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, *7*, 334–343.

Gouveia, R. R. dos R. (2018). *Construção de um sistema de controlo e performance numa startup*.

Guimarães, A. M. S. (2018). *Marketing digital e organizações: o contributo das redes sociais para as PME´ se as microempresas portuguesas*.

Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, *56*(3), 387–404. https://doi.org/10.2501/IJMR-2014-025

Hannula, M., & Pirttimaki, V. (2003). *Business intelligence empirical study on the top 50 Finnish companies*.

Hardie, W. (2019). *A Leader in 2019 Gartner Magic Quadrant for Operational DBMS*. Oracle Database Insider. https://blogs.oracle.com/database/post/oracle-a-leader-in-2019-gartner-magic-quadrant-for-operational-dbms

Hendrickse, J., Arpan, L. M., Clayton, R. B., & Ridgway, J. L. (2017). Instagram and college women's body image: Investigating the roles of appearance-related comparisons and intrasexual competition. *Computers in Human Behavior*, *74*, 92–100. https://doi.org/https://doi.org/10.1016/j.chb.2017.04.027

History.com Editors. (2021). Twitter Launches. *A&E Television Networks*. https://www.history.com/this-day-in-history/twitter-launches

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we instagram: A first analysis of instagram photo content and user types. *Eighth International AAAI Conference on Weblogs and Social Media*.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, *79*, 403–408. https://doi.org/https://doi.org/10.1016/j.procir.2019.02.106

HubSpot;mention. (2021). *2021 on Instagram: What do the numbers say?*

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, *37*(10), 751–768.

INE - Sistema de Contas Integradas das Empresas. (2021). *Empresas: total e por dimensão*. PORDATA.

https://www.pordata.pt/DB/Portugal/Ambiente+de+Consulta/Tabela

J. Clement. (2021). *Global top websites by monthly visits 2020*. STATISTA. https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/

Johnston, K., Chen, M.-M., & Hauman, M. (2013). *Use, Perception and Attitude of University Students Towards Facebook and Twitter*. https://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=c0038b73-4315-4c24-83d0-cd3e8417bcda%40sdc-v-sessmgr03

Jung, A.-R. (2017). The influence of perceived ad relevance on social media advertising: An empirical examination of a mediating role of privacy concern. *Computers in Human Behavior*, *70*, 303–309. https://doi.org/https://doi.org/10.1016/j.chb.2017.01.008

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review*, *28*(1), 21–37. https://doi.org/10.1016/S0363-8111(02)00108-X

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: the definitive guide to dimensional modeling*. Wiley.

Király, O., Potenza, M. N., Stein, D. J., King, D. L., Hodgins, D. C., Saunders, J. B., Griffiths, M. D., Gjoneska, B., Billieux, J., Brand, M., Abbott, M. W., Chamberlain, S. R., Corazza, O., Burkauskas, J., Sales, C. M. D., Montag, C., Lochner, C., Grünblatt, E., Wegmann, E., … Demetrovics, Z. (2020). Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. *Comprehensive Psychiatry*, *100*, 152180. https://doi.org/https://doi.org/10.1016/j.comppsych.2020.152180

Lauder, A. (2013). *Making connections: An evaluative study of the role of LinkedIn in alumni engagement*.

Linguateca. (2006). *Lists of Portuguese stopwords*. https://www.linguateca.pt/chave/stopwords/

Linvill, D. L., McGee, S. E., & Hicks, L. K. (2012). Colleges' and universities' use of Twitter: A content analysis. *Public Relations Review*, *38*(4), 636–638. https://doi.org/10.1016/j.pubrev.2012.05.010

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2010). *Risks and safety on the internet: the perspective of European children: key findings from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries*.

MacAvaney, S., Yao, H.-R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate speech detection: Challenges and solutions. *PLOS ONE*, *14*(8), e0221152.

Mathayomchan, B., & Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. *International Journal of Hospitality Management*, *90*, 102641.

https://doi.org/https://doi.org/10.1016/j.ijhm.2020.102641

Matos, A. P. M., Vieira, C. C., Amado, J., Pessoa, T., & Martins, M. J. D. (2018). Cyberbullying in Portuguese Schools: Prevalence and Characteristics. *Journal of School Violence*, *17*(1), 123–137. https://doi.org/10.1080/15388220.2016.1263796

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719.

Mccoy, C. G., Nelson, M. L., & Weigle, M. C. (2017). University Twitter Engagement: Using Twitter Followers to Rank Universities. *ArXiv*. www.timeshighereducation.com/world-university-rankings/2016/

McCracken, G. (1989). Who is the Celebrity Endorser? Cultural Foundations of the Endorsement Process. *Journal of Consumer Research*, *16*(3), 310. https://doi.org/10.1086/209217

Microsoft. (2021). *CALENDARAUTO function (DAX) - DAX*. Microsoft Docs. https://docs.microsoft.com/en-us/dax/calendarauto-function-dax

Ministerio da Administração Interna. (2019). *Relatório Anual de Segurança Interna 2019*. https://www.portugal.gov.pt/download-ficheiros/ficheiro.aspx?v=%3D%3DBQAAAB%2BLCAAAAAAABAAzNDA0sAAAQJ%2BleAUAAAA%3D

Moalla, I., Nabli, A., Bouzguenda, L., & Hammami, M. (2017). Data warehouse design approaches from social media: review and comparison. *Social Network Analysis and Mining*, *7*(1), 5.

Muñoz, A., Delgado, R., Rubio, E., Grilo, C., & Basto-Fernandes, V. (2017). Forum participation plugin for Moodle: Development and Discussion. *Procedia Computer Science*, *121*, 982–989.

Negash, S., & Gray, P. (2008). Business intelligence. In *Handbook on decision support systems 2* (pp. 175–193). Springer.

Neyman, C. J. (2017). *A survey of addictive software design*.

Nockleby, J. T. (2000). Hate speech in Encyclopedia of the American Constitution. *Electronic Journal of Academic and Special Librarianship*.

Nuțu, S.-C. P. (2019). The Battle for Words in the Digital Era. *Meridian Critic*, *32*(1), 137–144.

Obadă, R. (2019). *Sharing fake news about brands on social media: A new conceptual model based on flow theory*.

Olweus, D., & Limber, S. P. (2018). Some problems with cyberbullying research. *Current Opinion in Psychology*, *19*, 139–143. https://doi.org/https://doi.org/10.1016/j.copsyc.2017.04.012

Organisation for Economic Co-operation and Development. (2000). *Small Businesses, Job Creation and Growth: Facts, Obstacles and Best Practices*.

Orús, C., Barlés, M. J., Belanche, D., Casaló, L., Fraj, E., & Gurrea, R. (2016). The effects of learner-generated videos for YouTube on learning outcomes and satisfaction. *Computers & Education*, *95*, 254–269. https://doi.org/https://doi.org/10.1016/j.compedu.2016.01.007

Pantelimon, F.-V., Georgescu, T.-M., & Posedaru, B.-Ş. (2020). The Impact of Mobile e-Commerce on GDP: A Comparative Analysis between Romania and Germany and how Covid-19 Influences the e-Commerce Activity Worldwide. *Informatica Economica*, *24*(2), 27–41.

Peruta, A., & Shields, A. B. (2017). Social media in higher education: understanding how colleges and universities use Facebook. *Journal of Marketing for Higher Education*, *27*(1), 131–143. https://doi.org/10.1080/08841241.2016.1212451

Pfiffelmann, J., Dens, N., & Soulez, S. (2020). Personalized advertisements with integration of names and photographs: An eye-tracking experiment. *Journal of Business Research*, *111*, 196–207. https://doi.org/10.1016/j.jbusres.2019.08.017

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155–167. https://doi.org/https://doi.org/10.1016/j.chb.2016.03.084

Przybylski, A. K., Murayama, K., Dehaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, *29*(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Ribeiro, F., & Malheiro Da Silva, A. (2009). A Avaliação de Informação: uma operação metodológica. *IX Congresso ISKO Espanha e Portugal*. http://informationr.net/tdw/publ/papers/evaluation85.html

Richardson, J., Sallam, R., Kronz, A., & Sun, J. (2020). *Gartner Magic Quadrant for Analytics and Business Intelligence Platforms*. Gartner Research. https://www.gartner.com/en/documents/3980852/magic-quadrant-for-analytics-and-business-intelligence-p

Romiszowski, A. J. (2016). *Designing instructional systems: Decision making in course planning and curriculum design*. Routledge.

Salleh, S., Hashim, N. H., & Murphy, J. (2016). The role of information quality, visual appeal and information facilitation in restaurant selection intention. In *Information and Communication Technologies in Tourism 2016* (pp. 87–97). Springer.

Santos, C. L., Rita, P., & Guerreiro, J. (2018). Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. *International Journal of Educational Management*.

Saura, J. R. (2020). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation and Knowledge*. https://doi.org/10.1016/j.jik.2020.08.001

Shah, M., Cheng, M., & Fitzgerald, R. (2017). Closing the loop on student feedback: The case of Australian and Scottish universities. *Higher Education*, *74*(1), 115–129.

Shane-Simpson, C., Manago, A., Gaggi, N., & Gillespie-Lynch, K. (2018). Why do college students prefer Facebook, Twitter, or Instagram? Site affordances, tensions between privacy and self-expression, and implications for social capital. *Computers in Human Behavior*, *86*, 276–288. https://doi.org/https://doi.org/10.1016/j.chb.2018.04.041

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 1–9.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36.

Statista. (2021). *Twitter: most users by country* . https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Stuart, E., Stuart, D., & Thelwall, M. (2017). An investigation of the online presence of UK universities on Instagram. *Online Information Review*.

Talaue, G. M., AlSaad, A., AlRushaidan, N., AlHugail, A., & AlFahhad, S. (2018). The impact of social media on academic performance of selected college students. *International Journal of Advanced Information Technology*, *8*(4/5), 27–35.

The Commission of the European Communities. (2003). COMMISSION RECOMMENDATION of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32003H0361&from=PT

Times Higher Education (THE). (2021). *World University Rankings 2022*. Times Higher Education (THE). https://www.timeshighereducation.com/world-university-rankings/2022/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats

Troisi, O., Grimaldi, M., Loia, F., & Maione, G. (2018). Big data and sentiment analysis to highlight decision behaviours: A case study for student population. *Behaviour & Information Technology*, *37*(10–11), 1111–1128.

Twenge, J. M., & Martin, G. N. (2020). Gender differences in associations between digital media use and psychological well-being: Evidence from three large datasets. *Journal of Adolescence*, *79*, 91–102. https://doi.org/https://doi.org/10.1016/j.adolescence.2019.12.018

Twenge, J. M., Martin, G. N., & Spitzberg, B. H. (2019). Trends in US Adolescents' media use, 1976–2016: The rise of digital media, the decline of TV, and the (near) demise of print. *Psychology of Popular Media Culture*, *8*(4), 329.

Vannucci, A., Flannery, K. M., & Ohannessian, C. M. (2017). Social media use and

anxiety in emerging adults. *Journal of Affective Disorders*, *207*, 163–166. https://doi.org/https://doi.org/10.1016/j.jad.2016.08.040

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for ETL processes. *Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP*, 14–21.

Veletsianosa, G., Kimmonsb, R., Shawc, A., Pasquinid, L., And, & Woodwardb, S. (2017). Selective openness, branding, broadcasting, and promotion: Twitter use in Canada's public universities. *Educational Media International*. https://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=8bf9dbe7-2584-401b-ad01-c23ec678e9ca%40pdc-v-sessmgr06

W. H. Inmon. (2005). Building the Data Warehouse. In *Wiley Publishing*. Wiley Publishing. https://books.google.pt/books/about/Building_the_Data_Warehouse.html?id=QFK Tmh5IFS4C&printsec=frontcover&source=kp_read_button&hl=en&redir_esc=y# v=onepage&q=inman&f=false

Wang, J., Chen, T.-J., & Chiu, S.-H. (2005). Literature Review on Data Warehouse Development. *IACIS Pacific 2005 Conference Program*, 987–994.

Wang, Q., & Woo, H. L. (2008). The affordances of weblogs and discussion forums for learning: A comparative analysis. *Educational Technology*, 34–38.

Waszak, P. M., Kasprzycka-Waszak, W., & Kubanek, A. (2018). The spread of medical fake news in social media–the pilot quantitative study. *Health Policy and Technology*, *7*(2), 115–118.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, *1*.

Xu, A., Liu, H., Gou, L., Akkiraju, R., Mahmud, J., Sinha, V., Hu, Y., & Qiao, M. (2016). Predicting perceived brand personality with social media. *Tenth International AAAI Conference on Web and Social Media*.

Ye, Z., Hashim, N. H., Baghirov, F., & Murphy, J. (2018). Gender Differences in Instagram Hashtag Use. *Journal of Hospitality Marketing & Management*, *27*(4), 386–404. https://doi.org/10.1080/19368623.2018.1382415

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

Zhu, Q., Feng, Z., & Li, X. (2018). GraphBTM: Graph enhanced autoencoded variational inference for biterm topic model. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4663–4672.

Adrian, M., Feinberg, D., & Cook, H. (2019). *Gartner Magic Quadrant for Operational Database Management Systems*. Gartner Research. https://www.gartner.com/en/documents/3975492

Almeida, E., & Morais, N. S. (2020). The Use of Facebook and Instagram in Communication with New Students: the Case of Portuguese Public Universities. *SIIE*.

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79–95. https://doi.org/10.1007/s11747-019-00695-1

Arthurs, J., Drakopoulou, S., & Gandini, A. (2018). *Researching YouTube*. SAGE Publications Sage UK: London, England.

Avery, J., Steenburgh, T. J., Deighton, J., & Caravella, M. (2012). Adding Bricks to Clicks: Predicting the Patterns of Cross-Channel Elasticities over Time. *Journal of Marketing*, *76*(3), 96–111. https://doi.org/10.1509/jm.09.0081

Azeroual, O., & Theel, H. (2018). The Effects of Using Business Intelligence Systems on an Excellence Management and Decision-Making Process by Start-Up Companies: A Case Study. *International Journal of Management Science and Business Administration*, *4*(3), 30–40. https://doi.org/10.18775/IJMSBA.1849-5664-5419.2014.43.1004

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, *27*(2), 184–206.

Barbara Wixom, & Hugh Watson. (2010). *The Bi-Based organization*.

Biasutti, M. (2017). A comparative analysis of forums and wikis as tools for online collaborative learning. *Computers & Education*, *111*, 158–171. https://doi.org/https://doi.org/10.1016/j.compedu.2017.04.006

Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences*, *116*, 69–72. https://doi.org/https://doi.org/10.1016/j.paid.2017.04.039

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Burkhardt, J. M. (2017). *Combating fake news in the digital age* (Vol. 53, Issue 8). American Library Association.

Burt, R. S. (1999). The Social Capital of Opinion Leaders. *The ANNALS of the American Academy of Political and Social Science*, *566*(1), 37–54. https://doi.org/10.1177/000271629956600104

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod Record*, *26*(1), 65–74.

Cherry, B. (2018). *Why "Pages to Watch" Is a Valuable Facebook Marketing Tool*. Bluleadz. https://www.bluleadz.com/blog/pages-to-watch-tool-for-facebook-marketing

Chugh, R., & Ruhi, U. (2018). Social media in higher education: A literature review of Facebook. *Education and Information Technologies*, *23*(2), 605–616.

Ciechanowski, L., Jemielniak, D., & Gloor, P. A. (2020). TUTORIAL: AI research without coding: The art of fighting without fighting: Data science for qualitative researchers. *Journal of Business Research*, *117*, 322–330.

Constantinov, C., Popescu, P., & Mocanu, M. (2019). Identifying LinkedIn Usage Patterns Based on Feedback from Students and Fresh Graduates. *2019 18th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, 1–6. https://doi.org/10.1109/ROEDUNET.2019.8909507

Daniels, R. A., Pemble, S. D., Allen, D., Lain, G., & Miller, L. A. (2021). LinkedIn Blunders: A Mixed Method Study of College Students' Profiles. *Community College Journal of Research and Practice*, 1–16. https://doi.org/10.1080/10668926.2021.1944932

Dar, P. (2020). *Gartner's 2020 Magic Quadrant is Out*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/02/gartners-2020-magic-quadrant-bi-analytics-tools/

Davis, J., Wolff, H.-G., Forret, M. L., & Sullivan, S. E. (2020). Networking via LinkedIn: An examination of usage and career benefits. *Journal of Vocational Behavior*, *118*, 103396. https://doi.org/https://doi.org/10.1016/j.jvb.2020.103396

de Beer, D., & Matthee, M. (2021). *Approaches to Identify Fake News: A Systematic Literature Review BT - Integrated Science in Digital Age 2020* (T. Antipova (ed.); pp. 13–22). Springer International Publishing.

DeBettencourt, M. T., Williams, S. D., Vogel, E. K., & Awh, E. (2020). Sustained attention and spatial attention distinctly influence long-term memory encoding. *BioRxiv*, 2020.09.14.297341. https://doi.org/10.1101/2020.09.14.297341

Dell Technologies. (2018). *Digital Transformation Index*. Dell Technologies. https://www.delltechnologies.com/en-us/perspectives/digital-transformation-index.htm#scroll=off&pdf-overlay=//www.delltechnologies.com/asset/en-us/solutions/business-solutions/industry-market/dell_technologies_digital_transformation_index_ii_full_findings_report.pdf

DeSanctis, G., Fayard, A.-L., Roach, M., & Jiang, L. (2003). Learning in Online Forums. *European Management Journal*, *21*(5), 565–577. https://doi.org/https://doi.org/10.1016/S0263-2373(03)00106-3

Direção-Geral de Ensino Superior. (2021a). *1ª fase do Concurso Nacional de Acesso ao ensino superior público 2021 - Número de candidatos ao ensino superior volta a crescer e atinge novo máximo de cerca de 64 mil*. https://wwwcdn.dges.gov.pt/sites/default/files/nota_cna2021_1afase_v20agosto_2021.pdf

Direção-Geral de Ensino Superior. (2021b). *Ensino Superior em números*. https://www.dges.gov.pt/pt/pagina/ensino-superior-em-numeros

Directorate General for Higher Education. (2021). *Portuguese Higher Education System*. https://www.dges.gov.pt/en/pagina/portuguese-higher-education-system

Domenico, G. Di, Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341. https://doi.org/https://doi.org/10.1016/j.jbusres.2020.11.037

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52. https://doi.org/https://doi.org/10.1016/j.tsc.2013.12.004

ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., & Belding, E. (2018). Hate lingo: A target-based linguistic analysis of hate speech in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1).

Erdoğmuş, İ. E., & Çiçek, M. (2012). The Impact of Social Media Marketing on Brand Loyalty. *Procedia - Social and Behavioral Sciences*, *58*, 1353–1360. https://doi.org/https://doi.org/10.1016/j.sbspro.2012.09.1119

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. https://doi.org/https://doi.org/10.1016/j.copsyc.2019.03.005

Facebook Business Help Centre. (2021). *Reach | Facebook Business Help Centre*. Facebook for Business. https://www.facebook.com/business/help/710746785663278

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, *121*, 817–825.

Forret, M. L. (2018). *Networking as a job search and career management behavior. InU. C. Klehe & EW van Hooft*.

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, *51*(4), 1–30.

França, T., Cairns, D., Calvo, D. M., & de Azevedo, L. (2021). Lisbon, the Portuguese Erasmus city? Mis-match between representation in urban policies and international student experiences. *Journal of Urban Affairs*, 1–15. https://doi.org/10.1080/07352166.2021.1950544

GitHub. (2012). *Portuguese stop words*. https://gist.github.com/alopes/5358189

Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact. *International Journal of Market Research*, *59*(6), 703–705. https://doi.org/10.2501/IJMR-2017-050

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education*, *94*, 241–251.

Golfarelli, M., Maio, D., & Rizzi, S. (1998). Conceptual design of data warehouses from E/R schemes. *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, *7*, 334–343.

Gouveia, R. R. dos R. (2018). *Construção de um sistema de controlo e performance numa startup*.

Guimarães, A. M. S. (2018). *Marketing digital e organizações: o contributo das redes sociais para as PME´ se as microempresas portuguesas*.

Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, *56*(3), 387–404. https://doi.org/10.2501/IJMR-2014-025

Hannula, M., & Pirttimaki, V. (2003). *Business intelligence empirical study on the top 50 Finnish companies*.

Hardie, W. (2019). *A Leader in 2019 Gartner Magic Quadrant for Operational DBMS*. Oracle Database Insider. https://blogs.oracle.com/database/post/oracle-a-leader-in-2019-gartner-magic-quadrant-for-operational-dbms

Hendrickse, J., Arpan, L. M., Clayton, R. B., & Ridgway, J. L. (2017). Instagram and college women's body image: Investigating the roles of appearance-related comparisons and intrasexual competition. *Computers in Human Behavior*, *74*, 92–100. https://doi.org/https://doi.org/10.1016/j.chb.2017.04.027

History.com Editors. (2021). Twitter Launches. *A&E Television Networks*. https://www.history.com/this-day-in-history/twitter-launches

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we instagram: A first analysis of instagram photo content and user types. *Eighth International AAAI Conference on Weblogs and Social Media*.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, *79*, 403–408. https://doi.org/https://doi.org/10.1016/j.procir.2019.02.106

HubSpot;mention. (2021). *2021 on Instagram: What do the numbers say?*

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, *37*(10), 751–768.

INE - Sistema de Contas Integradas das Empresas. (2021). *Empresas: total e por dimensão*. PORDATA. https://www.pordata.pt/DB/Portugal/Ambiente+de+Consulta/Tabela

J. Clement. (2021). *Global top websites by monthly visits 2020*. STATISTA. https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/

Johnston, K., Chen, M.-M., & Hauman, M. (2013). *Use, Perception and Attitude of University Students Towards Facebook and Twitter*. https://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=c003 8b73-4315-4c24-83d0-cd3e8417bcda%40sdc-v-sessmgr03

Jung, A.-R. (2017). The influence of perceived ad relevance on social media advertising: An empirical examination of a mediating role of privacy concern. *Computers in Human Behavior*, *70*, 303–309. https://doi.org/https://doi.org/10.1016/j.chb.2017.01.008

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review*, *28*(1), 21–37. https://doi.org/10.1016/S0363-8111(02)00108-X

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: the definitive guide to dimensional modeling*. Wiley.

Király, O., Potenza, M. N., Stein, D. J., King, D. L., Hodgins, D. C., Saunders, J. B., Griffiths, M. D., Gjoneska, B., Billieux, J., Brand, M., Abbott, M. W., Chamberlain, S. R., Corazza, O., Burkauskas, J., Sales, C. M. D., Montag, C., Lochner, C., Grünblatt, E., Wegmann, E., … Demetrovics, Z. (2020). Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. *Comprehensive Psychiatry*, *100*, 152180. https://doi.org/https://doi.org/10.1016/j.comppsych.2020.152180

Lauder, A. (2013). *Making connections: An evaluative study of the role of LinkedIn in alumni engagement*.

Linguateca. (2006). *Lists of Portuguese stopwords*. https://www.linguateca.pt/chave/stopwords/

Linvill, D. L., McGee, S. E., & Hicks, L. K. (2012). Colleges' and universities' use of Twitter: A content analysis. *Public Relations Review*, *38*(4), 636–638. https://doi.org/10.1016/j.pubrev.2012.05.010

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2010). *Risks and safety on the internet: the perspective of European children: key findings from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries*.

MacAvaney, S., Yao, H.-R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate speech detection: Challenges and solutions. *PLOS ONE*,

14(8), e0221152.

Mathayomchan, B., & Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. *International Journal of Hospitality Management*, *90*, 102641. https://doi.org/https://doi.org/10.1016/j.ijhm.2020.102641

Matos, A. P. M., Vieira, C. C., Amado, J., Pessoa, T., & Martins, M. J. D. (2018). Cyberbullying in Portuguese Schools: Prevalence and Characteristics. *Journal of School Violence*, *17*(1), 123–137. https://doi.org/10.1080/15388220.2016.1263796

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719.

Mccoy, C. G., Nelson, M. L., & Weigle, M. C. (2017). University Twitter Engagement: Using Twitter Followers to Rank Universities. *ArXiv*. www.timeshighereducation.com/world-university-rankings/2016/

McCracken, G. (1989). Who is the Celebrity Endorser? Cultural Foundations of the Endorsement Process. *Journal of Consumer Research*, *16*(3), 310. https://doi.org/10.1086/209217

Microsoft. (2021). *CALENDARAUTO function (DAX) - DAX*. Microsoft Docs. https://docs.microsoft.com/en-us/dax/calendarauto-function-dax

Ministerio da Administração Interna. (2019). *Relatório Anual de Segurança Interna 2019*. https://www.portugal.gov.pt/download-ficheiros/ficheiro.aspx?v=%3D%3DBQAAAB%2BLCAAAAAAABA AzNDA0sAAAQJ%2BleAUAAAA%3D

Moalla, I., Nabli, A., Bouzguenda, L., & Hammami, M. (2017). Data warehouse design approaches from social media: review and comparison. *Social Network Analysis and Mining*, *7*(1), 5.

Muñoz, A., Delgado, R., Rubio, E., Grilo, C., & Basto-Fernandes, V. (2017). Forum participation plugin for Moodle: Development and Discussion. *Procedia Computer Science*, *121*, 982–989.

Negash, S., & Gray, P. (2008). Business intelligence. In *Handbook on decision support systems 2* (pp. 175–193). Springer.

Neyman, C. J. (2017). *A survey of addictive software design*.

Nockleby, J. T. (2000). Hate speech in Encyclopedia of the American Constitution. *Electronic Journal of Academic and Special Librarianship*.

Nuțu, S.-C. P. (2019). The Battle for Words in the Digital Era. *Meridian Critic*, *32*(1), 137–144.

Obadă, R. (2019). *Sharing fake news about brands on social media: A new conceptual model based on flow theory*.

Olweus, D., & Limber, S. P. (2018). Some problems with cyberbullying research. *Current Opinion in Psychology*, *19*, 139–143. https://doi.org/https://doi.org/10.1016/j.copsyc.2017.04.012

Organisation for Economic Co-operation and Development. (2000). *Small Businesses, Job Creation and Growth: Facts, Obstacles and Best Practices*.

Orús, C., Barlés, M. J., Belanche, D., Casaló, L., Fraj, E., & Gurrea, R. (2016). The effects of learner-generated videos for YouTube on learning outcomes and satisfaction. *Computers & Education*, *95*, 254–269. https://doi.org/https://doi.org/10.1016/j.compedu.2016.01.007

Pantelimon, F.-V., Georgescu, T.-M., & Posedaru, B.-Ș. (2020). The Impact of Mobile e-Commerce on GDP: A Comparative Analysis between Romania and Germany and how Covid-19 Influences the e-Commerce Activity Worldwide. *Informatica Economica*, *24*(2), 27–41.

Peruta, A., & Shields, A. B. (2017). Social media in higher education: understanding how colleges and universities use Facebook. *Journal of Marketing for Higher Education*, *27*(1), 131–143. https://doi.org/10.1080/08841241.2016.1212451

Pfiffelmann, J., Dens, N., & Soulez, S. (2020). Personalized advertisements with integration of names and photographs: An eye-tracking experiment. *Journal of Business Research*, *111*, 196–207. https://doi.org/10.1016/j.jbusres.2019.08.017

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155–167. https://doi.org/https://doi.org/10.1016/j.chb.2016.03.084

Przybylski, A. K., Murayama, K., Dehaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, *29*(4), 1841–1848.

https://doi.org/10.1016/j.chb.2013.02.014

Ribeiro, F., & Malheiro Da Silva, A. (2009). A Avaliação de Informação: uma operação metodológica. *IX Congresso ISKO Espanha e Portugal*. http://informationr.net/tdw/publ/papers/evaluation85.html

Richardson, J., Sallam, R., Kronz, A., & Sun, J. (2020). *Gartner Magic Quadrant for Analytics and Business Intelligence Platforms*. Gartner Research. https://www.gartner.com/en/documents/3980852/magic-quadrant-for-analytics-and-business-intelligence-p

Romiszowski, A. J. (2016). *Designing instructional systems: Decision making in course planning and curriculum design*. Routledge.

Salleh, S., Hashim, N. H., & Murphy, J. (2016). The role of information quality, visual appeal and information facilitation in restaurant selection intention. In *Information and Communication Technologies in Tourism 2016* (pp. 87–97). Springer.

Santos, C. L., Rita, P., & Guerreiro, J. (2018). Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. *International Journal of Educational Management*.

Saura, J. R. (2020). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation and Knowledge*. https://doi.org/10.1016/j.jik.2020.08.001

Shah, M., Cheng, M., & Fitzgerald, R. (2017). Closing the loop on student feedback: The case of Australian and Scottish universities. *Higher Education*, *74*(1), 115–129.

Shane-Simpson, C., Manago, A., Gaggi, N., & Gillespie-Lynch, K. (2018). Why do college students prefer Facebook, Twitter, or Instagram? Site affordances, tensions between privacy and self-expression, and implications for social capital. *Computers in Human Behavior*, *86*, 276–288. https://doi.org/https://doi.org/10.1016/j.chb.2018.04.041

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 1–9.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36.

Statista. (2021). *Twitter: most users by country* .

https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Stuart, E., Stuart, D., & Thelwall, M. (2017). An investigation of the online presence of UK universities on Instagram. *Online Information Review*.

Talaue, G. M., AlSaad, A., AlRushaidan, N., AlHugail, A., & AlFahhad, S. (2018). The impact of social media on academic performance of selected college students. *International Journal of Advanced Information Technology*, *8*(4/5), 27–35.

The Commission of the European Communities. (2003). COMMISSION RECOMMENDATION of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32003H0361&from=PT

Times Higher Education (THE). (2021). *World University Rankings 2022*. Times Higher Education (THE). https://www.timeshighereducation.com/world-university-rankings/2022/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats

Troisi, O., Grimaldi, M., Loia, F., & Maione, G. (2018). Big data and sentiment analysis to highlight decision behaviours: A case study for student population. *Behaviour & Information Technology*, *37*(10–11), 1111–1128.

Twenge, J. M., & Martin, G. N. (2020). Gender differences in associations between digital media use and psychological well-being: Evidence from three large datasets. *Journal of Adolescence*, *79*, 91–102. https://doi.org/https://doi.org/10.1016/j.adolescence.2019.12.018

Twenge, J. M., Martin, G. N., & Spitzberg, B. H. (2019). Trends in US Adolescents' media use, 1976–2016: The rise of digital media, the decline of TV, and the (near) demise of print. *Psychology of Popular Media Culture*, *8*(4), 329.

Vannucci, A., Flannery, K. M., & Ohannessian, C. M. (2017). Social media use and anxiety in emerging adults. *Journal of Affective Disorders*, *207*, 163–166. https://doi.org/https://doi.org/10.1016/j.jad.2016.08.040

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for ETL processes. *Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP*, 14–21.

Veletsianosa, G., Kimmonsb, R., Shawc, A., Pasquinid, L., And, & Woodwardb, S. (2017). Selective openness, branding, broadcasting, and promotion: Twitter use in Canada's public universities. *Educational Media International.* https://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=8bf9 dbe7-2584-401b-ad01-c23ec678e9ca%40pdc-v-sessmgr06

W. H. Inmon. (2005). Building the Data Warehouse. In *Wiley Publishing.* Wiley Publishing. https://books.google.pt/books/about/Building_the_Data_Warehouse.ht ml?id=QFKTmh5IFS4C&printsec=frontcover&source=kp_read_butto n&hl=en&redir_esc=y#v=onepage&q=inman&f=false

Wang, J., Chen, T.-J., & Chiu, S.-H. (2005). Literature Review on Data Warehouse Development. *IACIS Pacific 2005 Conference Program*, 987–994.

Wang, Q., & Woo, H. L. (2008). The affordances of weblogs and discussion forums for learning: A comparative analysis. *Educational Technology*, 34–38.

Waszak, P. M., Kasprzycka-Waszak, W., & Kubanek, A. (2018). The spread of medical fake news in social media–the pilot quantitative study. *Health Policy and Technology*, *7*(2), 115–118.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation.*

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, *1.*

Xu, A., Liu, H., Gou, L., Akkiraju, R., Mahmud, J., Sinha, V., Hu, Y., & Qiao, M. (2016). Predicting perceived brand personality with social media. *Tenth International AAAI Conference on Web and Social Media.*

Ye, Z., Hashim, N. H., Baghirov, F., & Murphy, J. (2018). Gender Differences in Instagram Hashtag Use. *Journal of Hospitality Marketing & Management*, *27*(4), 386–404. https://doi.org/10.1080/19368623.2018.1382415

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

Zhu, Q., Feng, Z., & Li, X. (2018). GraphBTM: Graph enhanced

autoencoded variational inference for biterm topic model. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4663–4672.

ANNEXES

ANNEX 1 - SQL Code of the tables in SQL Server Table

YouTube Query to retrieve videos with comments

```sql
select
        [Video URL]
from
        dbo.[ YoutubeVideosExtract]

where

        [Total comments] <> 0
```

FactComent_Load

```sql
/*Facebook Posts*/
Insert into FactComment ([PostID], [SourceID], [URL], [No Likes], [No
comments],[Post],[Publication Date], [No Shares],[User Name] )
        select
        [Post ID] as [PostID],
        '1' as [SourceID],
        [Link] as [URL],
        [Number of likes] as [No Likes],
        [Number of comments] as [No comments],
        [Post description] as [Post],
        [Post Date] as [Publication Date],
        [Number of Shares] as [No Shares],
        'Coimbra Business School' as [User Name]
        from
        FactFacebookPostsExtract

/*Google Maps Reviews*/
Insert into FactComment ([PostID], [SourceID],[Post],[Publication Date],[Review
Stars],[User Name])
        Select
        CONCAT([User Google acoount URL],[User Google picture],[user name],[time
of publish],[Stars],[review]) as [PostID]
        ,'8' as [SourceID]
        ,review as [Post]
        ,[time of publish] as [Publication Date]
        ,Stars as [Review Stars]
        ,[User Name]

        from [FactGoogleExtract]

/*Instagram*/
Insert into FactComment ([PostID], [SourceID],[Post],[Publication Date],[No
Likes],[No comments], [Reach],[User Name])
        Select
        CONCAT([Publication Date],[Reach],[No Likes],[No comments]) as [PostID],
        '2' as [SourceID],
        [Post],
        [Publication Date],
        [No likes],
        [No comments],
        reach,
        'Coimbra Business School' as [User Name]
        from [FactInstagramExtract]
/*Forum Coimbra*/
Insert into FactComment ([PostID], [SourceID],[Post],[Publication Date], [User
Name])
```

```sql
       Select
       CONCAT([User name],[User Forum Level],[date],[Comment]) as [PostID],
       '7' as [SourceID],
       Comment as [Post],
       [date] as [Publication Date]
       ,[User name]
       From [FactCoimbraForumExtract]


/*Uniarea*/
Insert into FactComment ([PostID], [SourceID],[Post],[Publication Date], [Topic
1], [Topic 2], [User Name] )
       Select
       Concat([Type],[Topic 1],[Topic 2],[Post no in
conversation],[comment],[user name],[Publish date]) as [PostID],
       '6' as [SourceID],
       comment as [Post],
       [Publish date] as [Publication Date],
       [Topic 1],
       [Topic 2],
       [user name]
       From [FactUniareaExtract]



/*Twitter*/
Insert into FactComment
([PostID], [Publication Date],[Post],[No Shares],[No Likes],[SourceID],[User
Name])
       select
       CONCAT([user name],[date],[tweet],[no of Retweets],[no of likes]) as
[PostID],
       [date] as [Publication Date],
       tweet as [Post],
       Cast([no of Retweets] as numeric) as [No Shares],
       Cast([no of likes] as numeric) as [No Likes],
       '4' as [SourceID],
       'Coimbra Business School' as [User Name]
       from
       FactTwitterExtract



/*Youtube*/
Insert into FactComment ([PostID], [SourceID], [URL], [No views] ,[No Likes], [No
comments],[Post],[Post Description],[Publication Date], [Duration],[User Name] )
       Select
       CONCAT([Video URL],[No of views],[No of likes],[No of dislikes],[No of
comments],[Video title],[Publication date],[Video duration]) as [PostID],
       '5' as [SourceID],
       [Video URL] as [URL],
       [no of views] as [No views],
       Cast([No of likes] as numeric) as [No Likes],
       Cast([No of comments] as numeric)as [No comments],
       [Video title] as [Post],
       [Video description] as [Post Description],
       [Publication date],
       [Video duration] as [Duration],
       'Coimbra Business School' as [User Name]
       from FactYoutubeExtract
```


DimUser_Load

```
/*Facebook*//*Instagram*//*LinkedIn*//*Twitter*//*Youtube*/
insert into DimUser ([UserID],[Name])
      values ('Coimbra Business School','Coimbra Business School')

/*Google Maps*/
insert into DimUser ([UserID],[Name], [No reviews], [User URL],[User Image])
      Select
      Concat([User Google acoount URL], [User Google picture],[user name]) as
      [UserID],
      [user name] as [Name],
      [User Google no reviews],
      [User Google acoount URL] as [User URL],
      [User Google picture] as [User Image]

      from [FactGoogleExtract]

/*Forum Coimbra*/
insert into DimUser ([UserID],[Name],[Level])
      Select distinct concat([user name],[User Forum Level]) as [UserID],
      [user name] as [Name],
      [User Forum Level] as [level]
      from FactCoimbraForumExtract

/*Uniarea*/
insert into DimUser ([UserID],[Name])
      Select distinct [user name] as [UserID],
      [user name]
      from FactUniareaExtract

/*Youtube Comments*/
insert into DimUser ([UserID],[Name],[User URL])
      select
      distinct (CONCAT([User URL],[Nickname])) as 'UserID'
      ,[Nickname] as [Name]
      ,[User URL]
      from
      (SELECT
            [Nickname]
            ,[User URL]
        FROM [Sentiment Analysis].[dbo].[FactYoutubeCommentsExtract])a
```

DimSource_Load

```
insert into DimSource ([SourceID], [name])
      values ('1','Facebook'),
      ('2','Instagram '),
      ('3','Linkedin'),
      ('4','Twitter'),
      ('5','Youtube '),
      ('6','Forum Uniarea'),
      ('7','Forum Coimbra'),
      ('8','Google')
```

Dim Term

```
Select

      postid,

      term
```

```sql
        FROM (

        SELECT postid,
        term,
        flag
FROM   (
          SELECT postid,
          upper(trim('""&:;0123456789/?|\-,.▪#"%ªº+*(!)-
@""''••^~." ' from [value])) AS term, (
           CASE
                WHEN [value] LIKE '%www.%' THEN 1
                WHEN [value] LIKE '%https%' THEN 1
                WHEN [value] LIKE '%HTTPS%' THEN 1
                WHEN [value] LIKE '@' THEN 1
                WHEN [value] IS NULL THEN 1
                WHEN [value]=' ' THEN 1
                WHEN [value]='' THEN 1
                WHEN [value] IS NULL THEN 1
                WHEN len([value])<2 THEN 1
                ELSE 2
                END ) flag
              FROM   (
                    SELECT      postid,
                                value
                    FROM        factcomment
                    CROSS apply string_split(post, ' ') AS bk) a ) b
WHERE  b.term NOT IN ( 'A', 'À', 'ACORDO', 'AGORA', 'AINDA', 'ALÉM', 'ALGUMAS',
'ALGUNS', 'ALTURA', 'ANO', 'ANOS', 'ANTES', 'ANTÓNIO', 'AO', 'AOS', 'APENAS',
'APESAR', 'APOIO', 'APÓS', 'AQUI', 'ÁREA', 'AS', 'ÀS', 'ASSIM', 'ASSOCIAÇÃO',
'ATÉ', 'ATRAVÉS', 'BANCO', 'BEM', 'CADA', 'CÂMARA', 'CAPITAL', 'CARLOS', 'CASA',
'CASO', 'CAUSA', 'CENTO', 'CENTRO', 'CERCA', 'CIDADE', 'CINCO', 'COM', 'COMISSÃO',
'COMO', 'CONSELHO', 'CONTA', 'CONTOS', 'CONTRA', 'CULTURA', 'DA', 'DAR', 'DAS',
'DE', 'DECISÃO', 'DEPOIS', 'DESDE', 'DESTA', 'DESTE', 'DIA', 'DIAS', 'DIRECÇÃO',
'DISSE', 'DIZ', 'DIZER', 'DO', 'DOIS', 'DOS', 'DUAS', 'DURANTE', 'E', 'É',
'ECONOMIA', 'ELE', 'ELEIÇÕES', 'ELES', 'EM', 'EMBORA', 'EMPRESA', 'EMPRESAS',
'ENQUANTO', 'ENTANTO', 'ENTÃO', 'ENTRE', 'EQUIPA', 'ERA', 'ESSA', 'ESSE', 'ESTA',
'ESTÁ', 'ESTADO', 'ESTADOS', 'ESTÃO', 'ESTAR', 'ESTAVA', 'ESTE', 'ESTES', 'EU',
'EUROPA', 'EUROPEIA', 'EXEMPLO', 'FACTO', 'FALTA', 'FAZ', 'FAZER', 'FERNANDO',
'FEZ', 'FIM', 'FINAL', 'FOI', 'FORA', 'FORAM', 'FORMA', 'FRENTE', 'GERAL',
'GOVERNO', 'GRANDE', 'GRANDES', 'GRUPO', 'GUERRA', 'HÁ', 'HISTÓRIA', 'HOJE',
'HOMEM', 'INÍCIO', 'INTERNACIONAL', 'ISSO', 'ISTO', 'JÁ', 'JOÃO', 'JOGO', 'JORGE',
'JOSÉ', 'LÁ', 'LADO', 'LEI', 'LHE', 'LISBOA', 'LOCAL', 'LUGAR', 'MAIOR', 'MAIORIA',
```

'MAIS', 'MANUEL', 'MAS', 'ME', 'MEIO', 'MELHOR', 'MENOS', 'MERCADO', 'MÊS', 'MESES', 'MESMA', 'MESMO', 'MIL', 'MILHÕES', 'MINISTÉRIO', 'MINISTRO', 'MOMENTO', 'MUITO', 'MUITOS', 'MUNDO', 'MÚSICA', 'NA', 'NACIONAL', 'NADA', 'NÃO', 'NAS', 'NEM', 'NESTE', 'NO', 'NOITE', 'NOME', 'NOS', 'NOVA', 'NOVO', 'NUM', 'NUMA', 'NÚMERO', 'NUNCA', 'O', 'OBRAS', 'ONDE', 'ONTEM', 'OS', 'OU', 'OUTRA', 'OUTRAS', 'OUTRO', 'OUTROS', 'P.', 'PAÍS', 'PAÍSES', 'PARA', 'PARECE', 'PARTE', 'PARTIDO', 'PARTIR', 'PASSADO', 'PELA', 'PELAS', 'PELO', 'PELOS', 'PESSOAS', 'PODE', 'PODER', 'PODERÁ', 'POLÍCIA', 'POLÍTICA', 'PONTOS', 'POR', 'PORQUE', 'PORTO', 'PORTUGAL', 'PORTUGUÊS', 'PORTUGUESA', 'PORTUGUESES', 'POSSÍVEL', 'POUCO', 'PRESIDENTE', 'PRIMEIRA', 'PRIMEIRO', 'PROBLEMA', 'PROBLEMAS', 'PROCESSO', 'PROGRAMA', 'PROJECTO', 'PRÓPRIO', 'PRÓXIMO', 'PS', 'PSD', 'PÚBLICO', 'QUAIS', 'QUAL', 'QUALQUER', 'QUANDO', 'QUANTO', 'QUASE', 'QUATRO', 'QUE', 'QUEM', 'QUER', 'QUESTÃO', 'R.', 'REGIÃO', 'RELAÇÃO', 'REPÚBLICA', 'SÃO', 'SE', 'SEGUNDA', 'SEGUNDO', 'SEGURANÇA', 'SEIS', 'SEJA', 'SEM', 'SEMANA', 'SEMPRE', 'SENTIDO', 'SER', 'SERÁ', 'SEU', 'SEUS', 'SIDO', 'SILVA', 'SISTEMA', 'SITUAÇÃO', 'SÓ', 'SOBRE', 'SOCIAL', 'SOCIEDADE', 'SUA', 'SUAS', 'TAL', 'TAMBÉM', 'TÃO', 'TARDE', 'TEM', 'TÊM', 'TEMPO', 'TER', 'TERÁ', 'TEVE', 'TINHA', 'TODA', 'TODAS', 'TODO', 'TODOS', 'TRABALHO', 'TRÊS', 'TUDO', 'ÚLTIMO', 'ÚLTIMOS', 'UM', 'UMA', 'VAI', 'VÃO', 'VER', 'VEZ', 'VEZES', 'VIDA', 'ZONA', 'DE ', 'A ', 'O ', 'QUE ', 'E ', 'DO ', 'DA ', 'EM ', 'UM ', 'PARA ', 'É ', 'COM ', 'NÃO ', 'UMA ', 'OS ', 'NO ', 'SE ', 'NA ', 'POR ', 'MAIS ', 'AS ', 'DOS ', 'COMO ', 'MAS ', 'FOI ', 'AO ', 'ELE ', 'DAS ', 'TEM ', 'À ', 'SEU ', 'SUA ', 'OU ', 'SER ', 'QUANDO ', 'MUITO ', 'HÁ ', 'NOS ', 'JÁ ', 'ESTÁ ', 'EU ', 'TAMBÉM ', 'SÓ ', 'PELO ', 'PELA ', 'ATÉ ', 'ISSO ', 'ELA ', 'ENTRE ', 'ERA ', 'DEPOIS ', 'SEM ', 'MESMO ', 'AOS ', 'TER ', 'SEUS ', 'QUEM ', 'NAS ', 'ME ', 'ESSE ', 'ELES ', 'ESTÃO ', 'VOCÊ ', 'TINHA ', 'FORAM ', 'ESSA ', 'NUM ', 'NEM ', 'SUAS ', 'MEU ', 'ÀS ', 'MINHA ', 'TÊM ', 'NUMA ', 'PELOS ', 'ELAS ', 'HAVIA ', 'SEJA ', 'QUAL ', 'SERÁ ', 'NÓS ', 'TENHO ', 'LHE ', 'DELES ', 'ESSAS ', 'ESSES ', 'PELAS ', 'ESTE ', 'FOSSE ', 'DELE ', 'TU ', 'TE ', 'VOCÊS ', 'VOS ', 'LHES ', 'MEUS ', 'MINHAS', 'TEU ', 'TUA', 'TEUS', 'TUAS', 'NOSSO ', 'NOSSA', 'NOSSOS', 'NOSSAS', 'DELA ', 'DELAS ', 'ESTA ', 'ESTES ', 'ESTAS ', 'AQUELE ', 'AQUELA ', 'AQUELES ', 'AQUELAS ', 'ISTO ', 'AQUILO ', 'ESTOU', 'ESTAMOS', 'ESTIVE', 'ESTEVE', 'ESTIVEMOS', 'ESTIVERAM', 'ESTÁVAMOS', 'ESTAVAM', 'ESTIVERA', 'ESTIVÉRAMOS', 'ESTEJA', 'ESTEJAMOS', 'ESTEJAM', 'ESTIVESSE', 'ESTIVÉSSEMOS', 'ESTIVESSEM', 'ESTIVER', 'ESTIVERMOS', 'ESTIVEREM', 'HEI', 'HAVEMOS', 'HÃO', 'HOUVE', 'HOUVEMOS', 'HOUVERAM', 'HOUVERA', 'HOUVÉRAMOS', 'HAJA', 'HAJAMOS', 'HAJAM', 'HOUVESSE', 'HOUVÉSSEMOS', 'HOUVESSEM', 'HOUVER', 'HOUVERMOS', 'HOUVEREM', 'HOUVEREI', 'HOUVERÁ', 'HOUVEREMOS', 'HOUVERÃO', 'HOUVERIA', 'HOUVERÍAMOS', 'HOUVERIAM', 'SOU', 'SOMOS', 'ÉRAMOS', 'ERAM', 'FUI', 'FOMOS', 'FÔRAMOS', 'SEJAMOS', 'SEJAM', 'FOSSE', 'FÔSSEMOS', 'FOSSEM', 'FOR', 'FORMOS', 'FOREM', 'SEREI', 'SEREMOS', 'SERÃO', 'SERIA', 'SERÍAMOS', 'SERIAM', 'TENHO', 'TEMOS', 'TÉM', 'TÍNHAMOS', 'TINHAM', 'TIVE', 'TIVEMOS', 'TIVERAM', 'TIVERA', 'TIVÉRAMOS', 'TENHA', 'TENHAMOS', 'TENHAM', 'TIVESSE', 'TIVÉSSEMOS', 'TIVESSEM',

```
'TIVER', 'TIVERMOS', 'TIVEREM', 'TEREI', 'TEREMOS', 'TERÃO', 'TERIA', 'TERÍAMOS',
'TERIAM',     'DE',     'DA',     'DO',     'COM',     'OS',     'AS',     'E')
AND    flag=2 ) a
```

ANNEX 2 - DAX codes in Power BI

DimCalendar

```
DimCalendar = CALENDARAUTO()
```

Month number

```
Month number = MONTH(DimCalendar[Date])
```

Month Name

```
Month Name =
SWITCH (
    DimCalendar[Month number],
    1, "Jan",
    2, "Feb",
    3, "Mar",
    4, "Apr",
    5, "May",
    6, "Jun",
    7, "Jul",
    8, "Aug",
    9, "Sep",
    10, "Oct",
    11, "Nov",
    12, "Dec"
)
```

Day

```
Day = DAY(DimCalendar[Date])
```

Year

```
Year = YEAR(DimCalendar[Date])
```

Weekday number

```
Weekday number = WEEKDAY(DimCalendar[Date], 1)
```

Week name

```
Weekday name =
SWITCH (
    DimCalendar[Weekday number],
    1, "Sunday",
    2, "Monday",
    3, "Tuesday",
    4, "Wednesday",
```

```
        5, "Thursday",
        6, "Friday",
        7, "Saturday",
        8, "Sunday"
    )
```

Company Size

```
Company Size =

SWITCH (

    DimLinkedInCompanySize[Company size],

    "1", "Small",

    "2-10", "Small",

    "11-50", "Small",

    "51-200", "Medium",

    "201-500", "Medium",

    "501-1000", "Large",

    "1001-5000", "Large",

    "5001-10000", "Large",

    "10001+", "Large",

    DimLinkedInCompanySize[Company size]

)
```

ANNEX 3 - Paper published in CISTI 2020

A análise de sensibilidade do Twitter numa Instituição de Ensino Superior utilizando Power BI

Twitter Sensitivity Analysis in a Higher School Using Power BI

Bruno Varela
Coimbra Business School | ISCAC, Polytechnic of Coimbra
Coimbra, Portugal
a2019117953@alumni.iscac.pt

Jorge Bernardino
Instituto Politécnico de Coimbra – ISEC
i2A – Instituto de Investigação Aplicada
Coimbra, Portugal
jorge@isec.pt

Isabel Pedrosa
Coimbra Business School | ISCAC, Polytechnic of Coimbra
Instituto Universitário de Lisboa (ISCTE-IUL) ISTAR-IUL, Portugal
ipedrosa@iscac.pt

Resumo — O Twitter, enquanto rede social, adquire especial importância nas instituições académicas, como é exemplo das Instituições de Ensino Superior (IES). Esta ferramenta permite às instituições, não só divulgar o seu trabalho e conhecer o *feedback* da comunidade quanto ao mesmo, mas também manter contacto com a sua rede *alumini* e fomentar conversações entre a comunidade académica. Neste trabalho elaborou-se um *dashboard* com os últimos 39 *Tweets*, relacionados com esta IES, com recurso à ferramenta de visualização de dados *Power BI*. Desta análise pode-se aferir que os utilizadores não utilizam o Twitter do com o intuito de debater ideias e colaborar, mas apenas para temas relacionados com atividades de lazer académico. Este *dashboard* permite perceber o que está a ser dito sobre esta IES no Twitter e de que forma se pode melhorar a utilização desta plataforma de comunicação.

Palavras Chave — *Twitter, Power BI, Análise de Redes Sociais, Comportamento Informacional, Gestão de Informação, Business Intelligence.*

Abstract - *Twitter, as a social network, acquires special importance in academic institutions, for example in Higher Education Institutions (HEI). This tool allows institutions to not only publicize their work and get feedback from the community about it, but also to keep in touch with their aluminum network and foster conversations between the academic community. In this work, a dashboard was created with the last 39 Tweets, related to this HEI, using the Power BI data visualization tool. From this analysis, we can assess users do not use Twitter do to debate ideas and collaborate, instead this platform is used mainly for topics related to academic leisure activities. This dashboard allows us to understand what is being said about this HEI on Twitter and how to improve the use of this communication platform.*

Keywords - *Twitter, Power BI, Social Media Analysis, Information Behavior, Information Management, Business Intelligence*

I. INTRODUÇÃO

Na presente Era de Informação em que vivemos, as instituições, seja qual for a sua natureza, necessitam de ter uma presença nas Redes Sociais, de forma a manterem-se competitivas. Contudo, nem sempre esta presença é concretizada da melhor forma, sendo bastante comum, o desalinhamento entre as publicações e a real estratégia da organização [1] ou a não utilização de todas as potencialidades destas ferramentas [2]. Assim o presente estudo tem como objetivo avaliar as iterações dos utilizadores no Twitter do de uma Instituição de Ensino Superior. Com este trabalho foi possível permitir a visualização de opiniões e sentimentos dos utilizadores, encontrar um padrão de utilização e elaborar sugestões de melhoria segundo os resultados. Para isto foram utilizadas a pesquisa documental [3] e a análise quantitativa com recurso à ferramenta de visualização de dados Power BI.

Assim o presente estudo é organizado da seguinte forma. Depois da introdução na segunda secção são abordados os trabalhos relacionados que possam servir de exemplo para o presente trabalho. Na terceira secção são descritas as metodologias aplicadas. Por sua vez na quarta secção, o desenvolvimento, são descritos todos os passos para o obter o Dashboard, sendo dividido em duas subsecções nomeadamente: *Developer membership no Twitter*, a aplicação do processo de *Extract, Transform, Load (ETL)*. A segunda subsecção está relacionada com o Desenvolvimento da Interface de Comunicação, ao nível das visualizações escolhidas e o seu layout. A quinta secção é leitura dos dados, onde são apresentados os resultados de forma quantitativa. A sexta secção, as discussões são apresentadas as relações generalizações apresentadas nos na secção anterior. Por fim na última secção, Conclusão e Trabalho Futuro, são descritos os principais resultados e que forma estes levam a um trabalho futuro.

II. TRABALHOS RELACIONADOS

São diversos os trabalhos de análise da utilização do Twitter, seja em conferências, ou em relatórios, contudo são poucas as publicações, em Business Intelligence, com a análise do Twitter em especial em Instituições de Ensino Superior. Assim, a título de contextualização, apresentam-se alguns trabalhos desenvolvidos na área.

Em [1] foi enfatizado que as Instituições académicas percebem a importância da utilização do Twitter para a difusão de informação, contudo ainda fazem um uso deficitário desta rede social. Por exemplo os *hashtags* utilizados eram sobretudo autorreferências, não sugerindo conversações mais amplas.

1

https://ieeexplore.ieee.org/document/9140979

Conferences > 2020 15th Iberian Conference ...

Twitter Sensitivity Analysis in a Higher School Using Power BI

Publisher: IEEE   Cite This   PDF

Bruno Varela ; Jorge Bernardino ; Isabel Pedrosa   **All Authors**

1 Paper Citation   **111** Full Text Views

Abstract

Abstract:
Twitter, as a social network, acquires special importance in academic institutions, for example in Higher Education Institutions (HEI). This tool allows institutions to not only publicize their work and get feedback from the community about it, but also to keep in touch with their aluminum network and foster conversations between the academic community. In this work, a dashboard was created with the last 39 Tweets, related to this HEI, using the Power BI data visualization tool. From this analysis, we can assess users do not use Twitter do to debate ideas and collaborate, instead this platform is used mainly for topics related to academic leisure activities. This dashboard allows us to understand what is being said about this HEI on Twitter and how to improve the use of this communication platform.

Authors

Citations

Keywords

Metrics

Published in: 2020 15th Iberian Conference on Information Systems and Technologies (CISTI)

Date of Conference: 24-27 June 2020

Date Added to IEEE *Xplore*: 15 July 2020

▶ **ISBN Information:**

Print on Demand(PoD) ISSN: 2166-0727

INSPEC Accession Number: 19786002

DOI: 10.23919/CISTI49556.2020.9140979

Publisher: IEEE

Conference Location: Seville, Spain

ANNEX 4 - Paper published in IC Marktech 2020

Chapter 10
The Application of Text Mining
for the Analysis of Connotation
in a Higher Education Institution

Bruno Varela, Jorge Bernardino, and Isabel Pedrosa

Abstract Twitter, as a social media, acquires special importance in academic institutions, such as promoting higher education institutions (HEIs). This tool allows not only to disseminate its work and know its feedback, but also to keep in touch with its alumni network and foster conversations between the academic communities. In this work, 118 tweets were extracted directly from Twitter, and a method was developed and applied that allows understanding the polarity of the tweets that mention a HEI, with three criteria: favourable, unfavourable and indifferent. In comparison with other methods, this model is an association of an individual assessment with a systematic assessment of terms, using linguistic tools. It is noteworthy that the method achieved about 67% of success.

10.1 Introduction

In the present Information Age in which we live, institutions, whatever their nature, need to have a presence on social networks to remain competitive. However, this presence is not always achieved in the best way, being quite common, the misalignment between publications and the real strategy of the organization [1] or the failure

B. Varela (✉) · I. Pedrosa
Coimbra Business School | ISCAC, Polytechnic of Coimbra, Polytechnic of Coimbra, Quinta
Agrícola – Bencanta, 3045-231 Coimbra, Portugal
e-mail: a2019117953@alumni.iscac.pt

I. Pedrosa
e-mail: ipedrosa@iscac.pt

J. Bernardino
Instituto Politécnico de Coimbra – ISEC, Coimbra, Portugal
e-mail: jorge@isec.pt

i2A – Instituto de Investigação Aplicada, Lagar dos Cortiços, 3045-093 Coimbra, Portugal

I. Pedrosa
ISCTE - Instituto Universitário de Lisboa, ISTAR-IUL, Avenida das Forças Armadas, 1649-026
Lisboa, Portugal

109

Published in Springer

Smart Innovation, Systems and Technologies, volume 205,
https://www.springer.com/series/8767

Springer Link

Smart Innovation, Systems and Technologies

Book Series
There are 235 volumes in this series
Published 2010 - 2022

About this series

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include