Article

Does User Generated Content Characterize Millennials’ Generation Behavior? Discussing the Relation between SNS and Open Innovation

Jose Ramon Saura 1, Felipe Debasa 2 and Ana Reyes-Menendez 1,*

1 Department of Business Economics, Rey Juan Carlos University, 28032 Madrid, Spain; joseramon.saura@urjc.es
2 Education Sciences, Language, Culture and Arts, Historical-Legal and Humanistic Sciences and Modern Languages, Rey Juan Carlos University, 28032 Madrid, Spain; felipe.debasa@urjc.es
* Correspondence: ana.reyes@urjc.es

Received: 12 August 2019; Accepted: 25 November 2019; Published: 26 November 2019

Abstract: The millennial generation plays a leading role in today’s connected world in which exists a confluence of numerous technologies and the internet in science, economy and innovation. This study aimed to identify the key factors that characterize the millennial generation within the online chatter on Twitter using an innovative approach. To this end, we analyzed the user generated content (UGC) in the social network (SNS) Twitter using a three-steps knowledge-based method for information management. In order to develop this method, we first used latent Dirichlet allocation (LDA), a state-of-the-art thematic modeling tool that works with Python, to analyze topics in our database. The data were collected by extracting tweets with the hashtag #Millennial, #Millennials and #MillennialGeneration on Twitter (n = 35,401 tweets). Secondly, sentiment analysis with a support vector machine (SVM) algorithm was also developed using machine-learning. Applying this method to the LDA results resulted in the categorization of the topics into those that contained negative, positive and neutral sentiments. Thirdly, in order to gather the final results, data text mining techniques were used. The negative factors that characterize the behavior of this generation are depression, loneliness and real-world relationship. The positive factors are body image, self-expression, travelers and digital life and the neutral factors are self-identity and anxiety. Practical implications can be used by public actors, companies or policy makers that are focused on the millennial generation as a target. The study has important theoretical applications as the topics discovered can be used to test quantitative models based on the findings and insights extracted from the UGC sample.

Keywords: user generated content; millennials behavior; data mining; SNS; open innovation

1. Introduction

In the first decade of the 21st century, the internet has caused innumerable changes at an economic, organizational and strategic level [1]. These changes brought about innovation initiatives in professional environments and have caused society to create new habits based on technology [2]. The study of these social changes has aroused the interest of researchers in recent years as well as brought considerable changes in the conceptualization of certain periods in human life and society [3]. In this way, scientific studies begin to focus their attention on the studies of social generations to learn to understand and empathize with them [1,3].

Likewise, starting from the generation Baby Boomers (1949–1968), youth started to be seen as a culmination of human growth [3]. Unlike their parents and grandparents, young people born in the 1960s and afterwards did not have to go to war and sought to live and enjoy life, experience new sensations, travel the world and acquire experiences from other cultures [4]. Furthermore,
a representative characteristic of the younger generations is freedom and the use of technology. This aspect is particularly noticeable in the generation of millennials (born 1981–1993). The millennials see freedom and the ability to make decisions based on their ideologies and identities as essential to their lives [4]. In this way, according to McDonal [5], the millennial generation understands life as a constant adventure. They do not want to accumulate material objects and value experiences instead. They do not want to settle down in a particular place and are global citizens, by analogy with the global digital dimension [4]. Therefore, the millennial generation is characterized by an extensive use of digital technologies and tools; for communication, they frequently use various social networks (SNS) [6] which are online platforms for the exchange of opinions and experiences that serve as a channel for communication and information which is called user generated content (UGC).

SNS users leave a digital trail in the form of data that can be analyzed [7]. While there are many SNS, in the present study, we focused on Twitter, which is one of the main SNS platforms that makes it possible to measure the opinion of millions of users, and therefore obtain an extensive sample of homogenous data [8] which is also massively used by millennials as presented by Godwin-Jones [7].

Extensive use of SNS generates dopamine and, therefore, can be considered addictive, such as drugs or gambling [9]. As shown by the number of centers that have emerged in recent years, the problem of technological addictions is a concern in many countries [10] and should be further analyzed as well as its relation to the business environment, global economy development and the influence of technology in the globalization phenomenon. The characterization of the millennial profile, both psychological and its abilities, has already been provided in several previous studies [11,12].

In the present study, we propose a new and original approach to the analysis of millennials based on the online chatter about this generation on Twitter, covering a research gap on the literature. Consequently, the primary purpose of this study is discovery, not hypothesis testing and not trying to control variables, but to discover them [13].

Therefore, the major aim of the present study is to establish a profile of the millennial generation users based on the online chatter in SNS, particularly Twitter, which has previously been reported to be a widely used medium for communication among this cohort [14].

SNS users generate large amounts of personal data stored on these platforms known as UGC [9]. Therefore, the public part of these data was used to develop a text-mining analysis approach. This approach was elaborated to find the main characteristics and behaviors of the millennials, including their concerns, motivations and abilities in order to be used by companies, social agents, policy makers and other interested public or private agents who have the millennial generation as their target audience in their business, marketing, social, cultural or sociological projects.

Several previous studies [14,15] demonstrated that it is possible to obtain key topics when analyzing the UGC in different industries and specific subjects and that the latent Dirichlet allocation (LDA) model can be meaningfully used to identify the topics in the UGC in SNS [15,16]. Therefore, the first research question addressed was as follows: Can the topics that psychologically characterize the millennial generation from the UGC on Twitter be identified? (RQ1).

Likewise, previous studies [17–19] employed various approaches to UGC content in order to identify the opinions and feelings in the comments and opinions of users on SNS such as Twitter, Google Maps, TripAdvisor, or Booking.com widely used by millennials [20]. In the present study, we focused on Twitter and sought to determine whether it was possible, through the analysis of the UGC on Twitter, to analyze user feelings (positive, negative and neutral) so the following research question was formulated: Can the analysis of the UGC on Twitter help identify the feelings associated with the topics that characterize the millennial generation? (RQ2).

Thirdly, the psychological characterization of the millennial generation, its beliefs and distinctive attitudes and behaviors are addressed in Doster [21] or Rodden and Hritz [22]. For instance, Ordun and Akun [11] studied attitudes and beliefs of the millennial generation. Furthermore, Sandfort and Haworth [14] identified the levels of emotional intelligence and personality characteristics of
the millennial generation. However, none of them proposed to obtain these indicators by analyzing SNS [23].

It is interesting therefore to investigate the indicators that characterize the main features of millennials that underlie their psychological profile on SNS [24] covering a research gap in the literature. According, the third research question addressed in this study was as follows: Can the analysis of the UGC on Twitter help identify the key psychological indicators and behaviors that characterize the millennial generation? (RQ3).

To reach our objectives, a three-phase methodology based on topic-modeling and data mining was applied for the first time to an UGC sample based on the online chatting about millennials on Twitter. First, an LDA model with Python was used, a state-of-the-art thematic modeling tool. The data were collected by extracting tweets with the hashtag #Millennials on Twitter ($n = 35,401$). Secondly, sentiment analysis was performed with a support vector machine (SVM) algorithm which is also used with machine learning. This made it possible to divide the identified topics into three sentiments, negative, positive, and neutral. Thirdly, textual analysis was further used with data text mining techniques. The application of this new approach based on UGC data to the millennial generation online chatting has proven important theoretical contribution in this study findings.

2. Literature Review

2.1. UGC Analysis

In this study, an online review or comment was defined as a piece of text that describes an experience about a specific product, service or topic generated by a SNS user in a public profile on the internet [25]. According to several researchers, by studying this type of UGC on the internet, a solid and powerful relationship can be built, thereby yielding robust results [26].

Nowadays, new technologies help to improve text data mining techniques and make it possible to automatically recognize the meaning of large databases or even topics within a database of comments, reviews or other types of content produced by SNS users [27]. A representative example of such technologies is the LDA technique which, in essence, is a modeling tool that can identify the topic of a database, such as qualitative reviews and comments, and quantify the number of mentions of a specific topic [28,29].

Of note, the analysis of comments made by users about other companies or users on the same SNS makes it possible to measure the total number of engagements. Nowadays, researchers are able to analyze users’ motivations to produce UGC [30,31]. For instance, Ordun and Akun [11] and Rodden and Hritz [22] have focused their research in UGC created by millennials to deepen the knowledge of psychological traits, attitudes and motivations [32]. Furthermore, Jia [29] and Monaco and Martin [33] identified major types of motivations of generating content on different millennial chat groups on the internet by analyzing all the chatting messages individually, i.e., without directly asking the users of those messages. The results indicated that the main motivations of these users were knowledge and sense of belonging [34].

Likewise, Reyes-Menendez et al. [35] investigated user motivations by studying their textual expressions on the internet. Another technique to improve the development of these methodologies is to estimate correlations between users and the rating number so that to obtain metrics related to the motivations and satisfactions of millennial internet users when generating content. For instance, in a study of millennial users’ feedback on TripAdvisor, Saura et al. [36] found that the behavior of the users can be understood through the study of their generated content. Owing to these UGC approaches, the companies can now understand the level of their consumers’ satisfaction with the products or services these companies offer [33,37]. For their part, Sandfort and Haworth [14] have deepened the understanding of millennials through the content generated in SNS to determine the personality traits of users and their degree of emotional intelligence.
Table 1 provides a summary of previous studies that used the UGC analysis.

Table 1. Characteristics of user generated content (UGC) analysis in previous research.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Rating</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Online Comments</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LDA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Social Interactions</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Topic Frequency</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: ✓ This research study contains the indicated UGC characteristic. Source: Adapted from Jia [29].

2.2. Sentiment Analysis for a Social Network Analysis

Sentiment analysis is a research methodology used to capture and analyze sentiments expressed in a given text sample [28]. The data are usually gathered in digital environments, such as online platforms or SNS. In previous research, sentiment analysis has proven to be able to effectively identify the expressed feelings, which makes this methodology particularly valuable for research on consumer decision making, attitudes and behavior [38].

Accordingly, several complementary methods and approaches have evolved to use sentiment analysis in a wide array of applications [28,39]. Some of these approaches are based on machine learning techniques and artificial intelligence, while other hybrid models are operated by specific software. Another alternative is training of algorithms along with data mining techniques to improve the probability of success of an algorithm operated with machine learning and to achieve more accurate results [38,40].

Furthermore, in several previous studies, models based on machine learning were developed for the analysis of different SNS, including online users’ opinions, or for the identification of the key factors related to a particular topic [39]. Supervised methods based on classification and categorization of key factors such as maximum entropy (MaxEnt) and SVM are widely used for SNS analysis [38]. To this end, one can use keywords, ratings of feelings regarding a topic, semantic meaning, concepts, semantic theories and sentiment-topic features such as hashtags, retweets or other valuations identifiers of any product and service on the internet [28,29].

For instance, Saura et al. [41] used a semi-supervised dual recurrent neural network and performed sentiment analysis of the content generated by SNS users. Furthermore, according to Jia [29], sentiment analysis is similar to traditional neural networks and can be used to evaluate a set of data over a long period of time. The author also indicated that this technique ensures a more effective and efficient sentiment analysis. Furthermore, in their study of the UGC with a hashtag (#WorldEnvironmentDay), Reyes-Menendez et al. [42] used sentiment analysis and found that managerial decision marking can be improved with the knowledge and insights acquired through UGC.

In summary, the results of previous body of work that used sentiment analysis suggest that sentiments in UGC can be analyzed using neural connections consisting of groups of users that interact in the context of UGC [39]. Secondly, textual analysis is the analysis of words and sentiments of online users and can help determine key factors and time. Thirdly, hashtags, URLs and mentions can be used to analyze specific groups of the users or specific topics [28]. Fourthly, a topic consists of the analysis of the sentiment based on the categories of the same content and, finally, the classification of the information regarding the sentimental analysis as sample keywords of information [43]. Table 2 shows the main characteristics of sentiment analysis.
Table 2. Previous studies that used sentiment analysis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuronal Connection</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Textual Analysis</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Time</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Hashtags, URLs or Mentions</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Classification of Information</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Source: Adapted from Saura et al. [41].

2.3. Textual Analysis

Textual analysis is a text mining analysis approach that determines the key factors in a big amount of data [13]. It is also a qualitative approach where, based on the frequency of specific items within a text, helps determine the keywords and sentiments in the analyzed content [38].

For instance, Vázquez and Escamilla [44] used a textual analysis with the NVivo software to identify the main attitudes and to determine the key factors to health among the elderly. Furthermore, Saito et al. [45] used textual analysis to predict the number of re-tweets based on the relevance of the USG content on Twitter. Saura et al. [41] develop a textual analysis process with the objective of identifying insights in a UGC sample collected from Twitter. In addition, the study uses qualitative processes to establish insights and identify knowledge about tweets published around the same topic. Also, Saura et al. [41] bases its analysis on the sample division by sentiments and nodes as well as the developed processes of data mining.

In this way, it can be concluded that textual analysis helps determine and identify the keywords of a greater relevance in a given sample and to investigate their impact by analyzing the content [28]. Therefore, textual analysis approaches have been widely used to prove that insights can be discovered from big amounts of data as presented by Arnold [32] and Reyes-Menendez [42].

Table 3 shows the main characteristics of the approximations based on textual analysis in order to identify the key factors in the UGC analysis.

Table 3. Main characteristics of textual analysis approaches.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification into Nodes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Categorization</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Word Count</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Keywords</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Source: Adapted from Saura et al. [8].

3. Methodology

The methodology used in this study was a three-phase process [28]. First, LDA with the Python model was used as a state-of-the-art thematic modeling tool to determine the topics in the database [29]. The data were collected by extracting tweets with the hashtag #Millennials, #Millennial and #MillennialGeneration on Twitter (n = 35,401) [41]. Secondly, sentiment analysis was performed with a SVM algorithm which was also used with machine learning in Python [36] that divided the identified topics into three sentiments [39]. Thirdly, textual analysis was further used with data text mining techniques, including the qualitative analysis software NVivo [46]. In order to summarize the methodology, Figure 1 shows the three-phase process used [28].
was based on a probabilistic assumption, assumed that content was generated in the following two steps: the first step required the identification of words and each word should be in an independent document leading to two consecutive steps [28]. These two steps consisted of randomly identifying the distributions of the topics identified in a sample and selecting the main topics found in that sample [29]. In real situations, it is difficult to estimate the distribution of topics over the document or the distribution of words over topics in advance [29].

After data collection, to start the proposed methods first, we developed an LDA model, which was based on a probabilistic assumption, assumed that content was generated in the following two steps: the first step required the identification of words and each word should be in an independent document leading to two consecutive steps [28]. These two steps consisted of randomly identifying the distributions of the topics identified in a sample and selecting the main topics found in that sample [29]. In real situations, it is difficult to estimate the distribution of topics over the document or the distribution of words over topics in advance [29].

After all the tweets were downloaded, some tweets were deleted due to repeated content, news, or retweets [15]. The images and multimedia files that were published along with the text of the tweet were excluded from the analysis [31]. Following Saura and Bennett [28], the sample of tweets was considered valid according to the following criteria: (i) active Twitter profiles, i.e., with activity during the three months after the use of the indicated hashtags; (ii) Twitter user profile with a profile photo and a cover photo; (iii) retweets from the same tweet using the indicated hashtags were removed, as such tweets were considered to be duplicate content; (iv) only public profiles and tweets in English using #Millennials, #Millennial and #MillennialGeneration between September and December 2018 following Saura et al. [41], Reyes-Menenez et al. [42] and Bologna and Hayashi [47]. As indicated in Saura and Bennett [28], the days of data collection in SNS can be established according to the parameters designed by the researchers. This procedure can be used if the main objective of the research is exploratory, rather than hypothetical testing and based on social media listening or social media analytics. In addition, for the collection of data, researchers must consider that there are no online events or movements on SNS during these days, within the industry studied, that may add noise and hinder the database.

Figure 1. Three-phase process in the proposed method. Source: The authors.
The importance between the hidden and observed variables is the joint distribution mathematically expressed as shown in Equation.

\[
ρ(β_1:k, θ_1:D, Z_{1:D}, ω_{1:D}) = \prod_{i=1}^{K} ρ(β_i) × \prod_{d=1}^{D} ρ(θ_d) × \sum_{n=1}^{N} ρ(Z_{d,n}|θ_d) \sum_{n=1}^{N} ρ(W_{d,n}|β_1:k, Z_{d,n}).
\]

- \(β_i\): Distribution of word in topic \(i\), altogether \(K\) topics.
- \(θ_d\): Proportions of topics in document \(d\), altogether \(D\) documents.
- \(z_{d,n}\): Topic assignment for the \(n\)th word in document \(d\), altogether \(N\) words.
- \(ω_{d}\): Observed words for document \(d\).

Consequently, the identification of the topics and words was conducted using Gibbs sampling using Equation (see Saura and Bennett [28] and Jia [29] for details of the process). Python software LDA 1.0.5 was used for the estimation.

Secondly, after the identification of the topics through the LDA model a sentiment analysis algorithm was used. This was done after training the machine with data-mining processes and subdividing the sample into positive, negative and neutral sentiments. Regarding the data mining development process, a total of 371 samples were collected. This allowed us to reach the average probability percentage of >0.723 Krippendorff’s alpha value (KAV) [13,46].

The basic formula for KAV is a relationship characterized by observed disagreement/expected disagreement. In Equation, an apparently simple proportion is shown, because the calculation method is computationally very complex. The calculation process involves resampling methods such as bootstrap [13,46]. The calculations in Equation are explained in depth in Krippendorff [13].

\[
\hat{α}'_K = \frac{P_a - P_e}{1 - P_e}, \quad \text{where} \quad \begin{cases} 
P_a = (1 - ε_a)P_a + ε_n, \quad ε_n = \frac{1}{2n} 
P_e = \sum_{k,l} W_{kl}P_{kl}, \quad P_{kl} = \sum_{k,l} W_{kl}P_{kl} \end{cases}
\]

This algorithm training was performed considering several factors, such as the identification of irony, sarcasm and contents related exclusively to the objective of this research. Throughout this process, the contents unrelated to #Millennials were excluded from the sample and training.

Thirdly, to start with the textual analysis approach, the different databases of feelings were processed by using NVivo, and the tweets were categorized into the following three nodes: Positive (\(N_1\)), Neutral (\(N_2\)) and Negative (\(N_3\)) [28,41]. The process of entering the data in NVivo was manual, although the databases were already divided based on the expressed type of sentiment. The next step was creating the structure of the nodes. Through filters, words such as connectors, prepositions or articles and their plural forms were eliminated [42].

The nodes were predefined data containers grouped according to their characteristics. It should be considered that the design and development of nodes is a conventional procedure used to analyze pure data and to achieve the highest possible descriptive and exploratory quality [5]. In this sense, a relevant indicator in the analysis with NVivo shows the weighted percentage [28,29]. This represents the weight of the indicators grouped into nodes based on the times words were repeated in the sample. To calculate the weighted percentage (WP), NVivo was used with the following formula (see Equation).
where \( K \) represents a query that allows the program to search the text. The behavior of each of the words and for each tweet can be seen. Therefore, the \( K \) value was found for the hashtag #Millennials. In this way, the average \( K \) for all the tweets was calculated in order to obtain the global value.

4. Analysis of Results

After applying the LDA model, the topics that characterize the millennial profile were identified (see Table 4). That is, the different words contained in the database were categorized into topics. In this respect, it should be noted that the names of the topics were defined automatically. This is a natural outcome when the LDA-based topic identification methodology is used [29]. In essence, the process of the development of the nomenclature of each topic is usually as follows. First, the most frequently used words mentioned by users are selected [29]. Second, phrases consisting of those most frequent words are formed [28]. Once the content categories make sense and form groupings of well-differentiated topics, the topic identification process is completed. Table 4 shows the topics identified in our UGC database.

<table>
<thead>
<tr>
<th>Topic Name</th>
<th>Topic Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Image</td>
<td>Body image topic includes the questions related to the cult of the body, the creation of content related to one’s image and the culture of concern about what other users think of their body image.</td>
</tr>
<tr>
<td>Depression</td>
<td>Depression topic includes the issues related to depression or bullying in digital environments.</td>
</tr>
<tr>
<td>Loneliness</td>
<td>Loneliness topic includes the contents related to the feeling of loneliness arising from the use of new technologies that create a parallel virtual reality.</td>
</tr>
<tr>
<td>Real-World Relationship</td>
<td>Real-world relationship topic includes the issues related to the concerns about the creation of personal relationships outside the internet, which is sometimes difficult due to the excessive use of social networks.</td>
</tr>
<tr>
<td>Self-Identity</td>
<td>Self-identity topic includes the contents related to the desire to create an identity of one’s own, both in the real world and on the internet.</td>
</tr>
<tr>
<td>Self-Expression</td>
<td>Self-expression topic includes the issues related to the need to find a digital channel to share the acts of self-expression. The issues reflect the need to show user self-identity.</td>
</tr>
<tr>
<td>Travelers</td>
<td>Travelers topic includes the issues related to travelling. It shows how the millennials share in social networks the contents about their trips and adventures.</td>
</tr>
<tr>
<td>Digital Life</td>
<td>Digital life topic includes the issues related to digital (online) life of users.</td>
</tr>
<tr>
<td>Startups</td>
<td>Startups topic includes the issues related to start-ups, technology-based companies in which millennials are comfortable because they can share their work life in the social network. Owing to the use of innovation and technology, such business models are more flexible.</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Anxiety topic includes the issues related to anxiety among users. Here, concerns include impatience and the need for an immediate reward.</td>
</tr>
</tbody>
</table>

In the next step, we identified the feelings associated with the identified topics. Importantly, the feelings associated with each topic are linked to the characteristics of the components of those topics. The average accuracy of the results of the application of sentiment analysis to each identified theme are presented following KAV reliability. KAV should obtain a result equal to or above 0.667, so that the results indicate that the algorithm has been trained a sufficient number of times, although Krippendorff [13] indicates that the minimum KAV should be adjusted according to the weight of the conclusions. In this sense, a KAV high could be ≥0.800 while a KAV between the measures 0.667–0.800 could be used to tentatively define and argue conclusions.
It should be noted that positive sentiments are the feelings related to positive values and connotations within the theme. Said differently, the fact that a subject obtains a positive feeling means that the factors that comprise this theme are perceived as positive by the millennials chatting on Twitter. For example, the topic of “Traveling” (see Table 5) gets a positive feeling, meaning that the millennials, according to their UGC on Twitter, perceive traveling, adventures and international activities in a positive way. The same principle applies to each of the identified topics. Table 5 shows the most frequent words in each topic and the identified sentiments and KAV.

Table 5. Sentiments according to each theme.

<table>
<thead>
<tr>
<th>Topic Name</th>
<th>Mean Words in the Topic</th>
<th>Sentiment</th>
<th>KAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Image</td>
<td>Body worship, sexual attractiveness, healthy body image, self-esteem, embodiment</td>
<td>Positive</td>
<td>0.791</td>
</tr>
<tr>
<td>Depression</td>
<td>Depressive disorder, being depressed, anger, irritability, helplessness, hopelessness</td>
<td>Negative</td>
<td>0.799</td>
</tr>
<tr>
<td>Loneliness</td>
<td>Loneliness, depression, suicide, bullying, sadness</td>
<td>Negative</td>
<td>0.715</td>
</tr>
<tr>
<td>Real-World Relationship</td>
<td>Virtual relationship, internet relationship, internet reality</td>
<td>Negative</td>
<td>0.801</td>
</tr>
<tr>
<td>Self-Identity</td>
<td>Oneself, beliefs, identity, social identity</td>
<td>Neutral</td>
<td>0.673</td>
</tr>
<tr>
<td>Self-Expression</td>
<td>Appearance, attitude, body language</td>
<td>Positive</td>
<td>0.735</td>
</tr>
<tr>
<td>Travelers</td>
<td>Travel, adventures, adrenaline, emotion, happiness</td>
<td>Positive</td>
<td>0.689</td>
</tr>
<tr>
<td>Digital Life</td>
<td>To be connected, social media, Instagram, YouTube, Facebook</td>
<td>Positive</td>
<td>0.669</td>
</tr>
<tr>
<td>Startups</td>
<td>Relaxed atmosphere, technology and innovation, casual atmosphere</td>
<td>Positive</td>
<td>0.783</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Instant reward, instantaneity, impatience, restlessness, ambition</td>
<td>Neutral</td>
<td>0.658</td>
</tr>
</tbody>
</table>

In the textual analysis step, we defined the factors or indicators that characterize each topic according to its sentiment (see Tables 6–8). To this end, we attended to both the psychological attitudes and behaviors linked to the sentiment of the identified topics. Specifically, using textual analysis with Nvivo software, three nodes were established to correspond to each topic according to their feeling, and based on these results, the text data mining was performed to determine the most important factors according to their weight within the selected topic. In this case, according to the number of times the words repeated in the dataset, they were classified in different nodes.

Table 6. Positive characterization of the millennial generation according to UGC chatting in node one (N₁).

<table>
<thead>
<tr>
<th>N₁</th>
<th>Positive Indicators</th>
<th>WP</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Image (BI)</td>
<td>The millennial generation feels the need to show their body to be accepted by their closest circle of friends. Healthy habits and healthy food consumption are necessary to feel the feeling of belonging to a group.</td>
<td>0.37</td>
<td>470</td>
</tr>
<tr>
<td>Self-Expression (SE)</td>
<td>The millennial generation openly shows their beliefs to their friends, even if these beliefs are not accepted. The millennial generation strives to generate their own social identity and feel belongingness in groups with similar ideologies.</td>
<td>0.29</td>
<td>319</td>
</tr>
<tr>
<td>Travelers (TR)</td>
<td>The millennial generation is an active generation, travels a lot and likes to share their experiences in social networks. The search for adrenaline and emotion is the basis of the adventure spirit of the millennial generation, above the price or the cost of the trip.</td>
<td>0.29</td>
<td>310</td>
</tr>
<tr>
<td>Startups (ST)</td>
<td>This topic encompasses the positive predisposition of the millennial generation to work in these companies characterized by labor flexibility and technology.</td>
<td>0.28</td>
<td>307</td>
</tr>
<tr>
<td>Digital Life (DL)</td>
<td>The millennial generation needs to feel that it is connected to the internet, it is their livelihood and they need to share their experiences through this medium. The main social networks used by the generation are, in the descending order of importance, YouTube, Instagram, Facebook and Twitter.</td>
<td>0.28</td>
<td>303</td>
</tr>
</tbody>
</table>
Once these similar words were grouped in the same independent nodes, a qualitative approximation was made to determine what the factor of each indicator was. Consequently, node one (N\textsubscript{1}) collected the factors related to positive sentiment, N\textsubscript{2} collected the factors related to neutral feelings and N\textsubscript{3} collected the factors related to negative feelings.

**Table 7.** Neutral characterization of the millennial generation according to UGC chatting in N\textsubscript{2}.

<table>
<thead>
<tr>
<th>N\textsubscript{2}</th>
<th>Neutral Indicators</th>
<th>WP</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Identity (SI)</td>
<td>The millennial generation needs to feel accepted by their followers on the internet. Social networks are the channel through which they constantly demonstrate their social identity.</td>
<td>0.30</td>
<td>313</td>
</tr>
<tr>
<td>Anxiety (AN)</td>
<td>The millennial generation seeks immediacy in achieving everything that is proposed. Impatience and the desire to achieve their goals quickly are factors that generate anxiety among this generation.</td>
<td>0.24</td>
<td>129</td>
</tr>
</tbody>
</table>

**Table 8.** Negative characterization of the millennial generation according to UGC chatting in N\textsubscript{3}.

<table>
<thead>
<tr>
<th>N\textsubscript{3}</th>
<th>Negative Indicators</th>
<th>WP</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression (DE)</td>
<td>Using the internet as a means to emotional expression usually causes a depressive disorder in this generation. This causes frustration and, when the millennials do not feel accepted in this digital environment, they feel unsupported, hopeless and lacking strength to continue with their daily tasks. Social acceptance in digital environments is the key factor leading to excessive use of social networks.</td>
<td>0.40</td>
<td>570</td>
</tr>
<tr>
<td>Loneliness (LO)</td>
<td>Excessive use of social networks makes the millennials feel alone in the real world. They create online communities. The feeling of loneliness is generated by two factors: the non-acceptance of the content published in social networks, and failure to achieve the expected success of engagement, (via followers) in social media</td>
<td>0.39</td>
<td>317</td>
</tr>
<tr>
<td>Real-World Relationship (RW)</td>
<td>The millennial generation has problems creating consolidated relationships. They are more comfortable in the digital environment, which hinders their relationships in the real world. Skills such as perseverance and hard work are sometimes lacking, as the millennials are accustomed to instant rewards. This frequently causes failures to establish lasting relationships in the real world.</td>
<td>0.26</td>
<td>291</td>
</tr>
</tbody>
</table>

5. Discussion

After analyzing the results, we have divided the topics that characterize millennials into the following two categories: topics related to psychological/personality traits and topics linked to attitudes and behaviors. It should be noted that this research has an exploratory based approach on the online chatting about millennials on Twitter. The insights presented as findings are grouped into the knowledge discovery science approaches that should be studied in future research to perform hypothesis testing and quantitative tests [28].

According to our results, millennials’ psychological/personality traits are positively determined by their body image (BI), a topic that demonstrates the concern of this cohort with their physical appearance. In this respect, our results are consistent with previous findings reported by Moon et al. [48] but diverge from the conclusions drawn by Doster [21] regarding the content shared by millennials about BI in SNS.

Furthermore, we found that the millennials are concerned with self-expression (SE), which leads them to frame their habits and abilities as parts of their ideologies and tastes through SNS. This result converges with the conclusion made by Foltz [49]. Overall, several previous studies have demonstrated
that, for millennials, important factors are the need to show the body to be accepted by the closest circle of friends as an act of self-expression, healthy habits (including healthy eating) linked to the need to show their body and the need to share positive emotions from their experiences (e.g., Grotts and Johnson [25]). Based on these findings, could we conclude then that SE is a predictor or consequence of BI in millennials? Our results linked both psychological/personality traits in this generation.

Regarding negative topics within psychological/personality traits, our results suggest that depression (DE), which is mainly caused by actions such as bullying in social networks, is underpinned by the excessive time this cohort spends in SNS. In this respect, our results support the conclusions of Chipps et al. [50] and Rogowski and Cohen [51] who identified a negative feeling for DE in millennials’ generation arising from SNS habits.

In the results, depression (DE) was linked with loneliness (LO), another negatively evaluated feature that we found. LO is elicited by the unlimited use of new technologies that create virtual realities as the main cause. This finding is congruent with the results previously reported by Morahan-Martin [52], although it has not yet been linked to the millennial generation as its characteristic feature [53].

Based on the results of our analysis, we concluded that these two negative descriptors, DE and LO, may lead to millennials’ inability to establish real-world relationships (RW). This inability to express feelings in a lasting way in time with people outside the internet and without an immediate reward has already been identified as a problem of future generations using the internet (e.g., McKenna et al. [27]). Yet, the immediate connection with the present-day generation of millennials has not been made yet.

With regard to neutral psychological/personality traits, our results demonstrated that the need for the feeling of self-identity (SI) drives millennials to create their own reality linked to their lifestyles and values that they share through SNS (see also Doster [21]). Therefore, self-expression is a typical characteristic feature of this cohort.

However, as demonstrated by Olson [54], energy and adventurous eagerness of this population group are counterbalanced by anxiety (AN) that arises due to the inability of getting instant reward. Interestingly, several other authors argue that anxiety in the millennial generation is linked not to instant rewards, but to their skills with technology at work (e.g., Reference [55]).

With regard to attitudes and behaviors that characterize millennials according to UGC online chatting, our findings suggested the following three topics: habits based on a digital life (DL); attitudes to travel (TR); and startups (ST) companies’ preferences to work.

Concerning the first topic, our results suggested that the most frequently used channels for UGC of the millennials’ chatting are YouTube, Instagram, Facebook and Twitter. These findings are important because each SNS contributes to the activation of certain psychological characteristics of individuals, such as concentration, intelligence, or patience [26].

With regard to the second topic, traveling, it does characterize the preferences of millennials, but has several downsides, such as the development of addiction to sharing information about one’s journeys [17,56]. This addiction is promoted by getting instant rewards from others, in the form of likes or comments on social platforms and can lead to generating addictions to social acceptance and the creation of parallel realities. Such additions can also have a negative impact on DE, LO or RW.

Finally, our results showed that positively evaluated topics in millennials’ UGC include startups preferences to work (ST). We linked this finding to the fact that work in startup promotes labor flexibility and allows millennials to better express their ideologies than in other types of professional environments [57]. This conclusion was also supported by Baum’s [58] findings.

In summary, we can conclude that using the internet as a dominant mean of self-expression defines the emotions experienced by millennials [34]. On the one hand, they can become increasingly irritable and, sometimes, their non-acceptance in digital environments generates a strongly negative affect with potentially devastating consequences, such as DE or LO. By contrast, millennials’ positive emotions are linked to their engagement in SNS communities and forms of social approval they receive there (such as likes, comments and the number of views of their actions on social platforms).
6. Conclusions

In the present study, we applied a three-phase methodology to analyze Twitter UGC to extract the main topics that, taken together, provide a thorough psychological characterization of the millennial generation and its behavior. We also identified the types of feelings associated with salient topics in this cohort. This made it possible to identify key factors within each of the identified topics.

The results contribute to the literature on psychological and behavioral profiling of different social and demographic groups, as well as to the research on UGC opinion [59]. Furthermore, we also propose a novel method, a psychological text-mining analysis technique that can be meaningfully used in further research to study other population cohorts and online communities as well as to topic discovery according to the online chatting on social media.

It has been demonstrated the method used is valid for topic and insights discovery as we have reached similar conclusions that other studies in this area using a similar method [14,60,61].

Regarding RQ1, it has been shown that UGC can be used to identify topics related to millennials as well as other topics [28,62] thus allowing the extraction of insights and generation of knowledge about the object of study of the investigation to be used in future investigations. The topics identified have also been classified into three sentiments and two categories, demonstrating that RQ2 proposed has been validated. The feelings identified in the topic related to the millennial generation can denote the importance that UGC communities bring to those. Therefore, based on these feelings, other research may focus on the specific analysis of expectations created by the UGC on negative or positive indicators shared through Twitter [63] as well as how users participate collaboratively [64].

Additionally, data mining techniques and textual analysis have been performed to obtain insights and knowledge from the database. These approaches have allowed us to answer positively to RQ3 as the insights obtained identify a psychological profile of the millennial generation according to the UGC chatting on Twitter and shows it behavior and habits [65,66].

As a result of these processes outlined, an important theoretical implication is the study of the 10 topics now directly linked to the study of millennial generation based on UGC that is the most significant contribution of this study (Positive topics: body image (BI), self-expression (SE), travelers (TR), startups (ST), digital life (DL); Neutral Topics: self-identity (SI), anxiety (AN) and Negative topics: depression (DE), loneliness (LO), real-world relationship (RW)).

If researchers take these insights as variables and constructs for their future models, they may be able to enhance their understanding of whether positive links exist between them by developing, for example, models based on partial least squares structural equation modeling (PLS-SEM) or Statistical Package for the Social Sciences (SPSS), analysis of moment structures (AMOS) among others, thus contributing to a field of research that emerges from approaches that extract knowledge from large amounts of UGC data.

Also, academics can use this research to better understand millennials main concerns and behavior to focus on the development of research within the field. In addition, they can focus on content analysis by users of different SNS to better understand what the main habits are when sharing information publicly in digital ecosystems and UGC.

From practical and managerial perspectives, our results are meaningful for industries that target millennials as their major consumer group. Managers and executives who have to develop business on the internet and digital marketing strategies can use our results to segment and personalize their advertising on SNS, if the target audience is millennials. In addition, managers and executives could consider our discoveries to develop content generation plans in SNS to increase the engagement with millennials in this digital environment.

Additionally, companies that develop projects in the field of sociology or the study of society in general, can use our results to establish patterns and models that characterize the millennial generation as a segment of the target market.
The limitations of the present study include a relatively small sample size and a limited time horizon. In further research, in order to properly address the concerns of this user group (including psychological problems and behavioral disorders), it would be necessary to determine other factors that could lead to a better understanding of the millennials as a generation. Likewise, one of the weaknesses that should be highlighted is that the diagnosis made about the millennial generation and its characterization and behavior analysis is the consequence of UGC chatting in social media and not by a clinical, social or in-depth interview method.

**Author Contributions:** J.R.S., A.R.-M. and F.D. conceived and designed the research study; J.R.S. performed the methodology; A.R.-M. and F.D. analyzed the results, J.R.S., A.R.-M. and F.D. wrote the paper.

**Funding:** This research received no funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


34. Antonacci, G.; Fronzetti Colladon, A.; Stefanini, A.; Gloor, P. It is rotating leaders who build the swarm: Social network determinants of growth for healthcare virtual communities of practice. *Int. J. Environ. Res. Public Health* 2013, 10, 261–272. [CrossRef]


58. Baum, T. A changing world of work. What can we learn from the service sector about employing Millennials (and Gen Z)? Organ. Dyn. 2019. [CrossRef]
63. Rossolatos, G. Negative brand meaning co-creation in social media brand communities: A laddering approach using NVivo. Psychol. Mark. 2019. [CrossRef]