

## Querying on Google Sheets

### Designing a Sentiments Analysis Alternative for rating tweets regarding the Ecuadorian 2021 Presidential Campaigns

Sariah López-Fierro<sup>1</sup> , Carlos Chiriboga<sup>2</sup>, and Rubén Pacheco<sup>3</sup>

<sup>1</sup> Universidad T. F. Santa María, Chile  
sariah.lopez@sansano.usm.cl

<sup>2</sup> Soluciones Wandarina S. A., Ecuador  
carlos@wandarina.com  
<http://www.wandarina.com>

<sup>3</sup> Universidad de E. Espíritu Santo, Ecuador  
rpachecov@uees.edu.ec

**Abstract.** This document contains an approach for the implementation of a sentiment analysis alternative after using Google Sheets for the rating of tweets retrieved with respect to the Ecuadorian Presidential 2021 Campaign.

**Keywords:** Google Clouds · Google Sheets · Querying · Data Analysis · Sentiment Analysis.

## 1 Introduction

Since 2004, when “Web 2.0” was loosely expressed and defined at a Silicon Valley Conference [1], software technologies worked on applications to improve the interaction and collaboration between the users, via the web [2].

During those years, Google not only managed to become the most popular search engine within the web [4], but it also evolved after introducing solutions such as Google News (2002), Gmail (2004), Google Maps (2005), among others; and expanding their business after buying popular web platforms such as Blogger (2003), Youtube (2006), Upstartle (2006), among others. Thus becoming in the “symbol of online innovation” [2].

In 2006, when Google Sheets was recently launched, some publications compared it to Microsoft’s office Excel. While its innovative online collaboration was highlighted, limitations in presentation, features, and formats were also noted [2] [5].

About 15 years after its release, Google Sheets is not only a collaborative sheet. Along the mathematical functions common in other spreadsheets, it also allows users to “program” or write custom functions through the Script Editor, which it is a JavaScript platform; and to “query” similar to SQL environments, which makes it capable of processing petabytes of data [3].

Being Google Sheets part of Google Cloud, thus being backed up by the power of High Performance Computers, this article aims to describe an alternative platform

for analysing data available for any user, while describing its implementation to process and interpret sentiment analysis from tweets retrieved during the Ecuadorian 2021 Campaigns.

## 2 Methodology: Querying on Google Sheets

### 2.1 Retrieving Tweets

To retrieve the tweets, we used Martin Hawksey's TAGS<sup>1</sup>. An available online script that allows you to easily set up a Twitter account and, through the Twitter API, collect the tweets. The only change that we made to his work was the addition of more fields in the "Archive" sheet that would allow us to obtain more details for a deeper analysis.

M. Hawksey's fields are: *id str, from user, text, created at, time, geo coordinates, user lang, in reply to user id str, in reply to screen name, from user id str, in reply to status id str, source, profile image url, user followers count, user friends count, user location, status url, entities str*.

Our fields were: *id str, text, status url, retweet count, favorite count, created at, user created at, in reply to screen name, user geo enabled, user location, place, in reply to user id str, in reply to status id str, user id str, user name, from user, user profile image url, entities str, lang, user description, user followers count, user friends count, user favourites count, user statuses count, source, extended tweet*.

### 2.2 Organizing Data

Google Sheets limits the use of its cells per document up to 5'000.000. Therefore, we were unable to work on the original TAGs file alone. Hence we established that each time we reached approximately the 100,000 row (100.000 tweets with details of the aforementioned fields), we would copy the data into a new spreadsheet.

Thus our step to organize the data involved to separate the tweets up to 100.000, in different files (from now on we will refer to them as **1-Archive**), also put them in different folders, consequently being able to work with them independently and then combine the results into a final step. Theoretically using the Divide-and-Conquer technique approach.

### 2.3 Cleaning Data

Our data was a mix of original tweets, retweets, and duplicate tweets that were separated into different files. As a result, in each one, we started by removing the duplicates through the *unique* command and then calling them to a new file (from now on we will refer to these new files as **2-Filter**).

Even though Google's "Sheets data connector for BigQuery" [3] would allow us to deal with large datasets "at once"; another way around to stick only in Google Sheets, was to use *importrange* command. Since this command works always with

<sup>1</sup> <https://tags.hawksey.info/get-tags/>

calls of 10.000 rows, we used it multiple times to import the 100.000+ rows from each *1-Archive* to fill the *2-Filter*.

Then, we separated the original tweets on a new sheet of the same document. From this new sheet, we used the *query* command.

### 2.4 Designing a Sentiment Analysis Alternative on Google Sheets

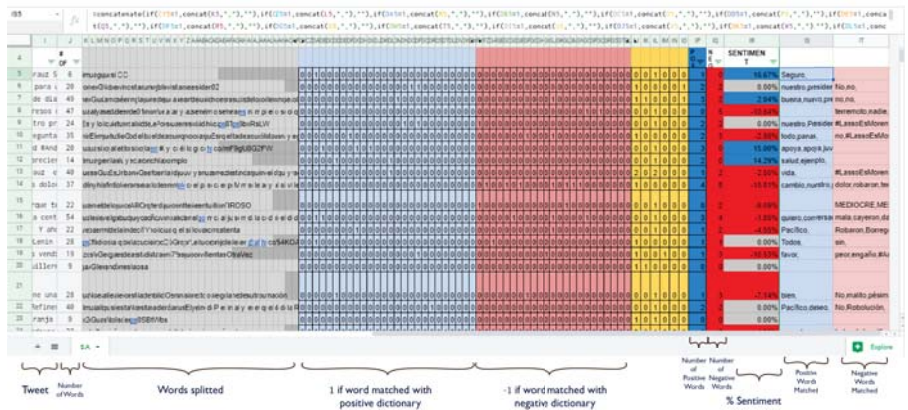


Fig. 1. Google Sheets - Screenshot of the Distribution of cells for analyzing tweets

Our approach for analyzing tweets in spreadsheets was to compare each word from each tweet against a dictionary of positive words and another of negative words. Add a positive value to each positive word found, or a negative value to each matching negative word, then calculate them. Besides, in order to get more accuracy in the rating, we decided to vary the value of the words from the dictionaries.

The final result determined the sentiment of each tweet. The higher the percentage of positive words, assures a positive tweet; and likewise, the lower the percentage of negative words identified a negative sentiment.

Besides, when a tweet had a result of “0”, we marked it as "Neutral" and it meant that a same number of positive and negative words were found in the tweet. Or that none of its words matched with any vocabulary included in the dictionaries.

**Setting up the dictionaries** We created a new file (from now on we will refer to it as *Dictionaries*) with two sheets. One sheet contained all the positive words, and the other the negative words. We only used one column from each sheet and filled them with all the respective words. As these dictionaries were made in Spanish, and we were about to match exact words, our dictionaries sought to include possible conjugations and variations with accent and without accent.



## 2.6 Limitations in the analysis

These are some of the limitations that we were able to find in our approach.

- To keep the document as small as possible (the fewer cells the better), we limited the analysis up to the 70th word of a tweet. If a tweet had more words than that, we ignored them.
- Ironies expressed with text or emojis were not well rated.
- If there was a tweet that replied an user, but mentioning a candidate, it was marked the sentiment towards the candidate.

## 3 Conclusion and Future Work

Analyzing data is mostly related to technologies that involved advanced technical support and constant rent in available public clouds. Our work included a different alternative that may respond to unattended parties who are more familiar with Google Suites.

Furthermore false-positives results allowed us to notice the limitations of measuring the sentiment with our approach. Thus for increasing the accuracy of analysing the tweets we have considered to measure the sentiment not only by words, but also in some cases by “phrases”; to determine sentiment according to context, to attempt to discriminate trolls from real accounts.

## Acknowledgement

This research has been supported and funded by the Technological Research Department of the company Soluciones Wandarina S. A. of Ecuador.

## References

1. Yu, C., and Du, H. (2007). Welcome to the World of Web 2.0. *The CPA Journal*, 77(5), 6.
2. Rienzo, T., and Han, B. (2009). Teaching Tip: Microsoft or Google Web 2.0 Tools for Course Management. *Journal of Information Systems Education*, 20(2), 123.
3. Gundrum, D. (2019, January 15). Connecting BigQuery and Google sheets to help with hefty data analysis. Retrieved March 15, 2021, from <https://cloud.google.com/blog/products/g-suite/connecting-bigquery-and-google-sheets-to-help-with-hefty-data-analysis>
4. Evans, M. P. (2007). Analysing Google rankings through search engine optimization data. *Internet research*.
5. Firth, M., and Mesureur, G. (2010). Innovative uses for Google Docs in a university. *Jalt call journal*, 6(1), 3-16