Tweeting Power:

Analyzing sentiment and rhetoric by Colombian Congressional Members

William Lowe

Thesis Supervisor

María José de Jesús Lee Ocampo

Master of Data Science for Public Policy

2024

Word Count: 6175

Hertie School of Public Policy

Abstract

This paper analyzes the language used by members of Colombia's *Cámara de Representantes* on Twitter during the initial three months of their terms, focusing on how sentiment, content, and public reactions reflect the evolving political alignments and gender dynamics. It observes a slow but ongoing reduction in the gender gap, alongside a significant leftward political shift among traditionally right-leaning parties.

The sentiment analysis reveals a general trend of positivity in tweets, even when discussing controversial political figures. Public engagement analysis indicates a preference for negative content, as evidenced by a negative correlation between tweet sentiment and the number of likes received, suggesting complex dynamics in public interaction with political communication.

Table of Contents

Introduction

How[1] humans communicate between themselves, but also to powerful figures, has changed dramatically over the past three centuries. This paper aims to shed light on the language used by members of Colombia's Cámara de Representantes on Twitter during the first three months following their appointment.

Previous research on the topic within the Colombian context has predominantly focused on the general trends in political communication, without delving into the specifics of language use, or the reactions this generated. This study seeks to contribute to this literature by providing a focused analysis of sentiment, content, and public engagement with tweets. Additionally, this study aims to expand upon political alignments and gender dynamics within the chamber.

The method, as well as the findings, can be divided into three main groups. First, on a general note, the gender gap in Congress, though steadily closing, still shows a clear preference for Male representatives. On the political side, a significant shift has been made towards the left by many traditionally right-leaning parties, though it is ultimately unclear if this change in allegiance will hold.

On the Sentiment Analysis side, the main takeaway seems to be a strong overall preference, on the side of the members of Congress, for writing tweets that could be classified as positive, additionally, this trend seems to hold even when talking about potentially controversial figures, such as the current, as well as former Presidents. However, it is unclear if this is a genuine trend or a result of difficulty in getting the entire corpus of mentions of said political figures.

Finally, on the side of the public reaction to the tweets, the number of likes a particular tweet received is negatively related to the associated score said tweet received,

[1] Link to the Github Repository is: https://github.com/LeeMariaj/Thesis---Twetting-Power

showcasing a preference in viewers for negative tweets, which did not seem to react to the gender of the congress member or the affiliated party. Lastly, some conclusions, as well as possible research avenues are discussed.

Literature Review

The relationship between authority figures and ordinary citizens has shifted significantly over the last three centuries, going from the late 18th century, when all countries were effectively closed autocracies, to the 21st century, where half of the world's countries are democracies, and even then, within the autocratic governments, at least half of them still have some sort of electoral process (Herre et al., 2024). Today, a defining feature of democracies is the ability to hold their leaders accountable for their actions and decisions (Schmitter, 2007).

Just as governmental power in and of itself has evolved over these last centuries, so have communication services. This period has seen people go from communicating by letters to phone calls, to emails, and finally, by instant messages and social media posts.

As human life has begun to rely more on the use of technology and social media, so have politicians jumped at the opportunity to have a direct line to their citizens (Buccoliero et al., 2020). It is in this particular context that we see microblogging platforms, such as Twitter, rise as a tool used by politicians, as well as other actors, to influence election results (Buccoliero et al., 2020), as well as communicate progress or setbacks on their work once elected (Ceron, 2017).

In response to this, social researchers have also turned their attention to social media, and in particular, Twitter, to study a wide array of phenomena, from everyday interactions to the language and behavior exhibited through key historical events that have happened over

the past two and a half decades, such as economic collapses, global pandemics, national uprisings, war, etc.

Academics have argued that within the Colombian context, Twitter[2] has established itself as a source of communication between citizens and the political elite, as well as a meeting point for discussion during times of political unrest (Gallego et al., 2019).

Examples of this include an analysis of the effect that bots focused on de-polarizing discussions during the final stage of the Peace Plebiscite of 2016 could have on voting (Gallego et al., 2019). As well as an analysis of the vocabulary utilized on Twitter to discuss the treaty negotiation from various state and non-state actors (Barreto-Galeano et al., 2021).

Other studies have instead chosen to analyze the words political leaders within Colombia used to communicate during the COVID-19 pandemic, finding that former President, Iván Duque[3]'s 50 most popular tweets were mostly of either informative or patriotic nature, meanwhile, at the time Congress Member and Current Colombian President, Gustavo Petro[4]'s 50 most popular tweets were mostly critiques to the handling of the pandemic by the Government, amongst other topics (Haman et al., 2022).

Research Question

While the ultimate goal would be to understand how Congressional Members from Colombia communicate with the public, given the gap between the scope of that question and the resources associated with the writing of this Master's thesis, this paper will instead concentrate on analyzing the language that the Members of Colombia's *Cámara de*

[2] As of January 2022, Twitter boasted an active user base of around 5.6 million Colombians (Rodríguez, 2023).
[3] Presidential term 2018 to 2022.
[4] Presidential term 2022 to 2026.

Representantes[5] for the period 2022 to 2026 used on their Twitter, posts from the first three months since they were appointed (July 20, 2022 to October 20, 2022).[6]

The choice to sample members from the *Camara de Representantes*, instead of from the *Senado de la República*, or potentially even from both, stemmed from the logic and manner by which both Chambers are formed. While the latter holds 117 seats (*Conoce a todos los congresistas 2022 - 2026 de Colombia*, n.d.) which are all elected at the national level from a common pool, every *circunscripción territorial*[7] is alloted at least two seats in the *Cámara de Representantes* (with additional seats being allotted according to population size), meaning this chamber is significantly more likely to hold a more diverse pool of representatives, and hence, reflect different political and communication styles.

It is also worth noting that only a part of the entire congressional sample is being analyzed from the *Cámara de Representantes*, namely, 138[8] members from the 194 elected (which constitutes approximately 71.1% of the members).

Because communication is a wide and abstract concept, this paper has chosen to concentrate on a select number of aspects. From some more surface-level, such as the gender of the communicator (in our case the members of Congress), the length of it (in our case of the tweets), the number of times they choose to communicate, and the choice by the author of said communication to complement their text with additional resources such as images, videos, or links to news articles.

---

[5] House of Representatives of Colombia, alongside the *Senado de la República* (Senate of the Republic) makes up the Legislative Branch for the Colombian Government (*Cámara De Representantes*, n.d.).

[6] This particular range was chosen as it would allow for a certain personal style to develop for each congress member, without allowing for too much refinement by a potential social media manager.

[7] Territorial constituency, originally referring to every *departamento*, or state, amongst which Colombia is divided, as well as the Capital City, however, additional constituencies can be added to assure representation of marginal groups (*Cámara De Representantes: Mapa Electoral*, n.d.).

[8] Of which, 24 were found to either have no twitter account, or have no available tweets for the period of interest.

Following this, the focus shifts to content, with every tweet being analyzed by a sentiment analysis model which assigns it a score between -1 and 1, showcasing how "positive" or "negative" it is. Additionally, each tweet is searched for mentions of the name of Colombia's Current President, Gustavo Petro, as well as the past three. This finding was then further complemented by the predicted emotion the previous step assigned to each tweet.

While Colombia has made significant advances in terms of gender equality[9], these visible advances have been, as concluded by a study conducted by Colombia's *Departamento Administrativo Nacional de Estadistica*[10], *Consejería Presidencial para la Equidad de la Mujer*[11] and *ONU Mujeres*[12], slow, biased and ultimately insufficient (Merino & López Barajas, n.d.). This, complimented by the fact that this election cycle marks the highest proportion of women elected to serve within Congress (Botero, 2022; Merino & López Barajas, n.d.), justifies the need to carry out a gender-based analysis.

On the other, as it stands, this particular administration constitutes a momentous point in Colombia's history, as it coincides with the election of its first left-leaning President, Gustavo Petro. Unfortunately, this has been accompanied by a high level of turmoil, controversy, and name-calling from both sides of the political spectrum. Given this context, results will also be separated by the political association of the respective Congress members that make them.

On this point, an acclaratory point is necessary, given that congress members are divided not by political association, but by political parties, this would be the natural division along which to analyze the data. However, this is complicated by the high number of political parties that are currently active within the *Camara de Representantes* (upwards of 30).

---

[9] Such as ruling that at least a third of all congressional candidates from each party must be from each gender (*Ley 581 de 2000*, 2000).
[10] National Administrative Department of Statistics.
[11] Presidential Advisory for Women's Equity.
[12] UN Women.

Because of the above, it was decided to group each political party by whether it had declared itself[13] to stand with the government, with the opposition[14], or if it had instead claimed to be independent. An additional "Peace" category was then included to account for members within the *Cámara de Representantes* who were given their seats as part of an effort to guarantee representation to otherwise neglected communities (*Cámara De Representantes: Mapa Electoral*, n.d.) To carry out this analysis, several primary sources were consulted (Bonilla, 2022; Colombia Renaciente, 2022; Infobae, 2022; La Silla Vacía, 2024; Muñoz, 2022; Redacción Política, 2022).

Finally, as communication is ultimately a two-way street, this paper also seeks to showcase the relation between the variables mentioned above and the number of likes and retweets that they generate, as this can be considered as a direct show of engagement and popularity of particular content for the members.

Methods

To carry out this analysis, the first step was to go through *Congreso Visible*[15]'s webpage, made available by *Universidad de los Andes*, to have a complete list of all Congress Members that belonged to *Camra de Representantes*, as well as their respective full names, their Twitter accounts, their political party, and their gender (*Conoce a todos los congresistas 2022 - 2026 de Colombia*, n.d.).

From there, using the advanced search tool from Twitter, the posts made by each Congress member within the relevant period were displayed, allowing for the manual collection of each of them within a dedicated Excel file. In addition to the textual content of the tweet, information regarding the date it was posted, the number of likes, the number of

[13] Within colombia, political organizations are obligated within the first month of a new presidency to declare themselves as either Opposition, Independent or a Government Organization (Ley 1909 de 2018, 2018).
[14] Worth noting is that, as a political party, declaring oneself as part of the opposition is not only a protected right, but is also accompanied with additional funding, access to the communication services of the state (such as radio), among other rights (Ley 1909 de 2018, 2018).
[15] Visible Congress.

retweets, whether there was a photo accompanying the tweet[16], whether there was a video accompanying the tweet[17], and if there was a link to a news article was also collected. Once the collection process was finalized, there were a total of 21.506 tweets from 114 congress members.

To carry out the Sentiment Analysis, some considerations had to be taken into account. The first of these considerations was. naturally, how exactly to go about it. Sentiment Analysis is in general understood as the process by which categorization is given to a particular grammatical unit (word, sentence, paragraph, etc.), such that it can be labeled as either showcasing a particular sentiment, or a particular intensity of a particular sentiment (Alsaeedi & Khan, 2019; Giachanou & Crestani, 2017; Hutto & Gilbert, 2014).

To achieve this, there seem to be at least 3 methods, i) establishing some sort of baseline, and manually categorizing each grammatical unit by either one or multiple participants, ii) developing a significant corpus of pre-labeled data, either with specific terms or entire grammatical units, train a machine learning model with this corpus, and then ask it to predict over the data you wish to categorize, or iii) using a language learning model which has already been trained by another researcher, and applying it to your particular case.

After significant research, a decision was made to utilize the VADER[18] language model, developed by C.J. Hutto and Eric Gilbert. This decision was made given that, while in general language learning models seem to have difficulties assigning a particular sentiment to tweets as they are relatively short and are often composed of abbreviations or otherwise informal language (Giachanou & Crestani, 2017), VADER was shown to work particularly well for social media media, were it was found that "(it) even outperform(ed) individual

[16]For the image category, a distinction was made depending on wether the image contained the politician's face, wether it contained informative text, wether it was an invitation on the side of the member to participate in a particular event, or wether the image fit none of the above mentioned criteria.

[17] For the video category, a distinction was made depending on wether the video showed the face of the politician or not.

[18] for Valence Aware Dictionary for sEntiment Reasoning.

human raters (...) at correctly classifying the sentiment of tweets into positive, neutral, or negative classes." (Hutto & Gilbert, 2014).

Additionally, all tweets whose scores were found to be exactly equal to 0, or which had no text to analyze, were discarded. The first of these decisions was made under the impression that a score of exactly 0, could be a sign that the model was unable to generate a classification based on the language utilized, and hence, this would ultimately have introduced bias within the model, as a lack of ability by the model to recognize sentiment does not mean a lack of sentiment within a tweet. This step ultimately lead to the reduction of the sample from 21.506 tweets to 21.441 tweets.

However, one important caveat associated with applying VADER to this particular case was its lack of support for languages outside of English. This, given it was trained with English sentences and furthered fine-tuned by using English-specific heuristics, to carry out its evaluation (Hutto & Gilbert, 2014). To work around this, it was necessary to then translate each tweet before giving it to VADER for classification.

To do this, the tool deep_translator, developed by Nidhal Baccouri, was utilized to pass each tweet into Google Translate (2020), after which said translation was given to the VADER language model, where the composite score[19] was then saved and interpreted as reflecting the sentiment of the tweet. After which, a posterior classification was made between "positive" and "negative" tweets, where, as suggested by the author, scores above 0.05 are considered to reflect a positive tweet, scores below -0.05 are considered to reflect a negative tweet, and, scores between -0.05 and 0.05 are considered to reflect a neutral tweet.

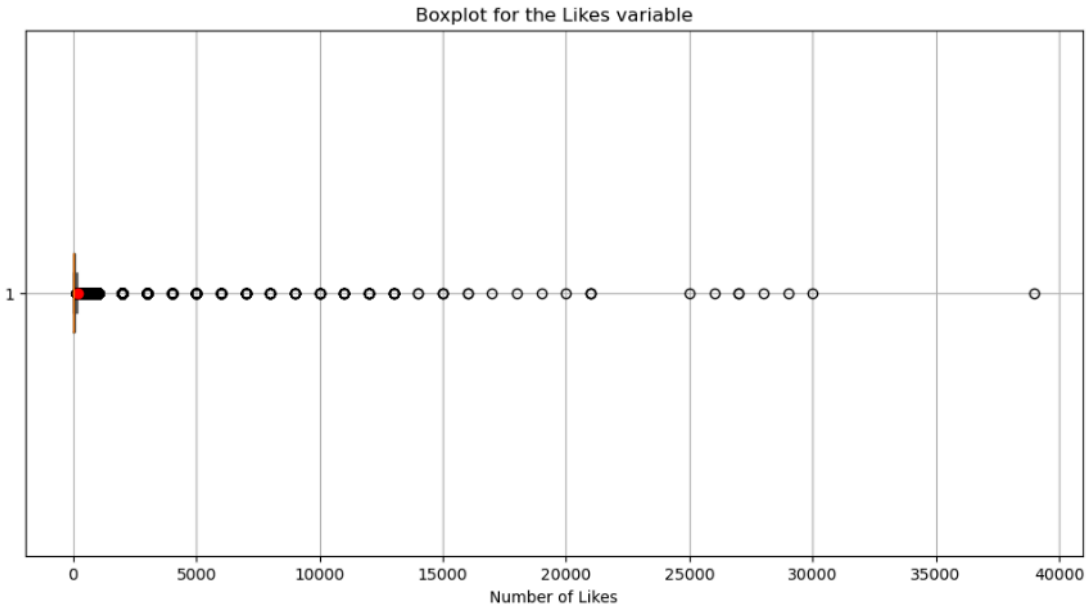
A drawback of this particular approach is the possibility of an intrinsic and essential part of the text that is associated with the way it was written originally being lost because of the translating process. This is of course a real possibility, however, because the nature of

[19] Calculated by summing the valence score of each word according to its previous classification, adjusting said sum to reflect common heuristics, and then normalizing it to fit between -1 and 1 (Hutto & Gilbert, 2014).

sentiment analysis, is usually not to carefully consider the particular intention behind a particular grammar unit, but instead, to try and generate a more objective measure, according to the words chosen, as well as the grammatical signs and other additions, the author believes that, that which could potentially be lost by translating, would have likely not have been picked up by the sentiment analysis model in the first place.

Once the sentiment analysis was finalized, the resulting data was first cleaned, and then utilized to calculate certain basic statistics, as well as run simple regression models to try and further conceptualize the relation between variables such as the sentiment score associated with a particular tweet with the number of likes it received.

Figure 1

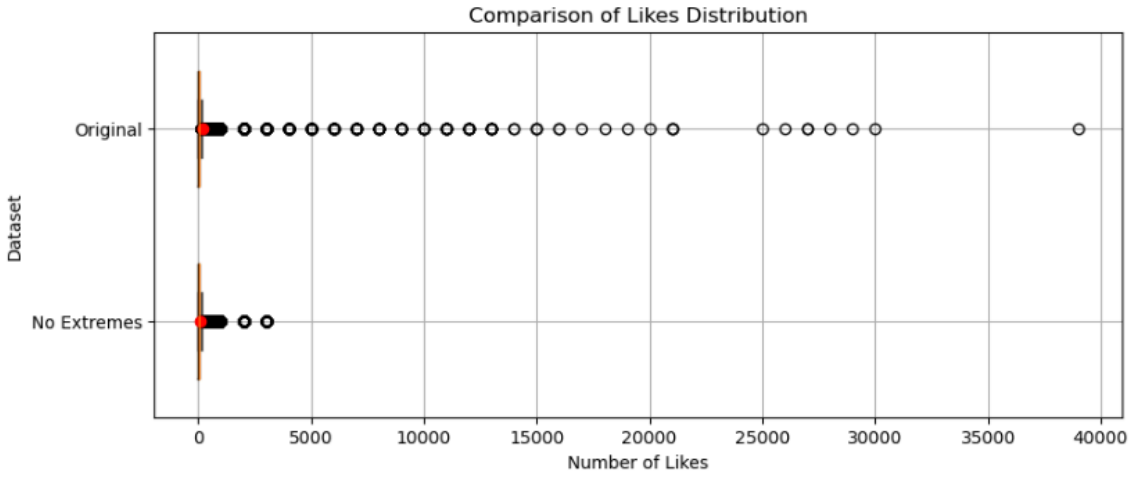


It was at this stage that it was discovered that the variable holding the number of likes that each tweet had received was severely skewed thanks to a significant number of outliers, to the point where the resulting boxplot, as can be seen in Figure 1, despite holding close to the entirety of the sample, is reduced to barely more than a line.

To deal with this situation, the z-score of the number of likes for each tweet was calculated, and any tweet that was found to have likes that were more than 3 standard

deviations away from the mean was subsequently prevented from being taken into account when analyzing the number of likes with other relevant variables. The distribution of the modified dataset can be seen in Figure 2.

Figure 2



Results

General Statistics

The analysis is in general divided into three levels, i) for the *Camara de Representantes* as a whole, ii) by the gender of the members, and iii) by the political association of the party the members belong to. Table 1 then refers to the dataset in its most general terms.

Table 1

Number of Members	**Number of Tweets**	**Avg. number of Tweets per Member**	**Avg. number of words per Tweet**
114	21440	188.07	29.71

Table 2, then, seeks to quantify the difference between members according to their gender. Here, one of the most noteworthy results is the fact that of the total number of tweets within the dataset, only around 27.44% were made by women. Also worth noting is that

despite being less represented within Congress, and therefore the database, they still write more tweets on average than their male counterparts, despite both writing tweets of roughly the same length.

Table 2

Gender	Number of Members	Number of Tweets	Avg. number of Tweets per Member	Avg. number of words per Tweet
Men	84	15557	185.2	29.48
Women	30	5883	196.1	30.33

Lastly, congress members were divided according to the stated allegiance of their respective party, or the nature of the seat that they hold, which is in turn reflected in Table 3. Here we see that the database is largely dominated, both in terms of members and in terms of tweets, by parties associated with the government. This reflects the dominance that the political party currently in charge (*Pacto Historico*[20]) had at the beginning of its period in terms of other political party allies (Galvis, 2022), as many traditionally right-leaning parties chose to ally themselves with the current administration.

Table 3

Party Association	Number of Members	Number of Tweets	Avg. number of Tweets per Member	Avg. number of words per Tweet
Government	78	15234	195.31	30.96
Independent	12	1981	165.08	24.19
Opposition	14	3557	254.07	26.77
Peace	9	480	53.33	32.41

Particularly interesting amongst this abundance of support for the current administration, is the significant number of tweets made by members of the opposition on

[20] Historical Pact.

average, compared to the same metric for either the Independent Association as well as the Peace Association, which while slightly smaller in terms of members are vastly surpassed in both the absolute number of tweets, but even more acutely, in the average number of tweets metric[21].

Gender by Association

Figure 3

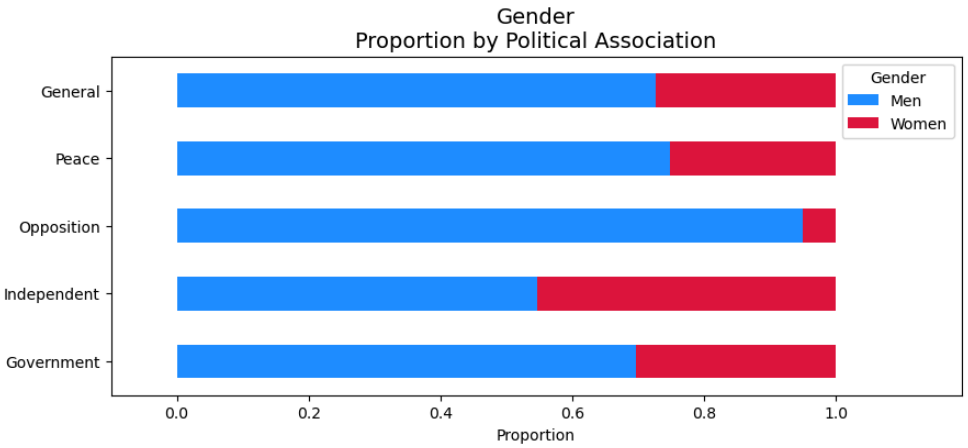


Figure 3, provides additional insights into the quite disparaging situation of gender representation within Congress. In general, the gender proportion of our sample is very close to the actual gender proportion within the *Camara de Representantes*, with the sample placing this number at 26.32%, as opposed to the actual number, which is 28.35%.

Worth noting is the considerably low proportion of women members of *Camara* that belong to the Opposition, though this number is significantly lower than the actual number[22], it remains the lowest proportion out of the sample. This statistic is further soured given that

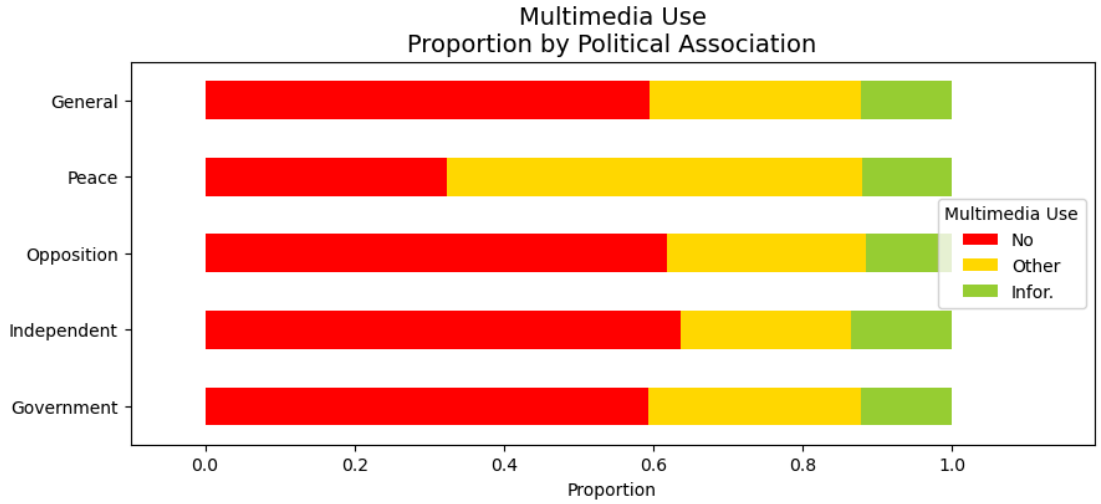
[21] A partial explanation of this phenomenon might relate to the fact that, given the shift in political sentiminet showcased by the election of the first left leaning president in Colombia's history, the members of the right that managed to hold on to their seats in power, were, in turn, some of the more politically established. Meaning having both the resources, as well as the political interest to produce a significant number of tweets.

[22] The actual proportion is closer to 21%, as out of the 19 congress member that belong to the Opposition, 4 are women (Galvis, 2022).

only the Independent association is made up of at least ⅓ of women, meaning that while parties may be indeed meeting the requirements provided by law, there is still significant work to be done in this respect.

Use of Multimedia Resources by Association

Figure 4



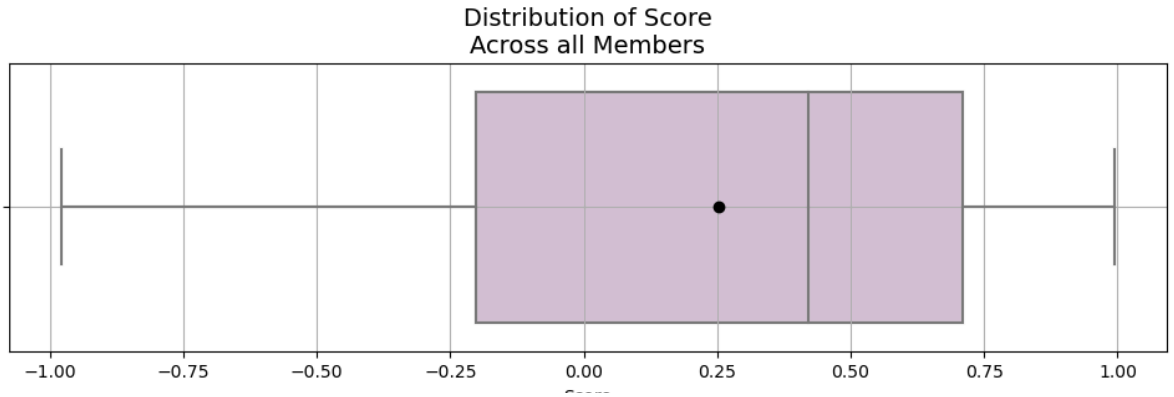
The use of multimedia is understood here as making a tweet that is accompanied by either an image, a video, or a link to a news article. Additionally, the distinction between informative and other refers to an internal classification within the database. If the tweet was accompanied by either a news article or an image that contained text of informative character, then it would be considered "Informative". On the other hand, images of congress members, political cartoons, images commemorating a particular date, videos[23], etc, are classified as "Other".

[23]This means that some potentially informative videos are here being classified as "other", though by their nature they should instead be classified as "informative". This is of course, far from ideal, however, given the fact that trying to further classify a video as informative or not, would not only have added a significant amount of time to the data collection process, but also would have opened this classification to a significant level of bias.

Overall, we see that the behavior regarding utilizing informative media seems to be pretty consistent amongst all 4 party associations. The only significant deviation seems to be within the Peace Association, which uses about double the amount of multimedia resources within their tweets, though this effect is largely dominated by the "other" category.

Negative and Positive Speech

Figure 5

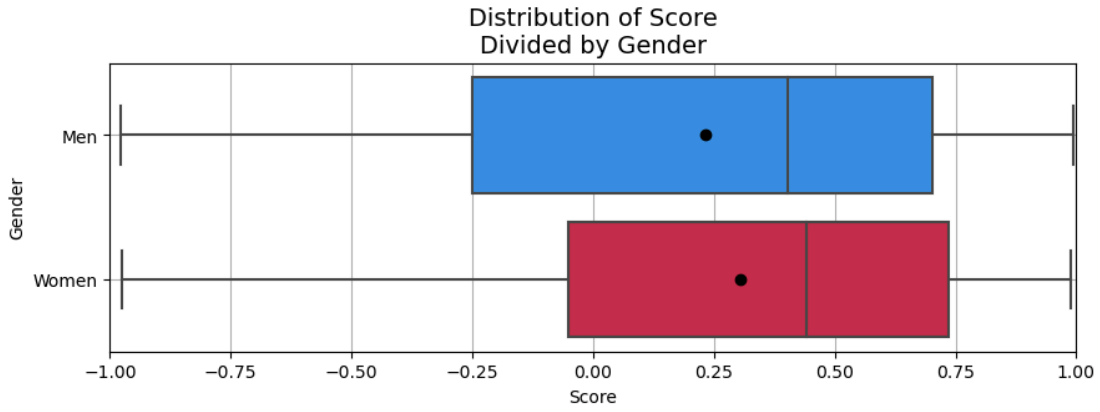


In general, we observe that in our overall sample, there is ample representation for the entire range of possible scores, however, there seems to be a preference for positive results. This is further emphasized by the fact that the overall proportion of tweets classified as positive (57.61%) more than doubles the proportion of tweets classified as negative (21.44%). This trend can be further confirmed by Figure 5.

By Gender

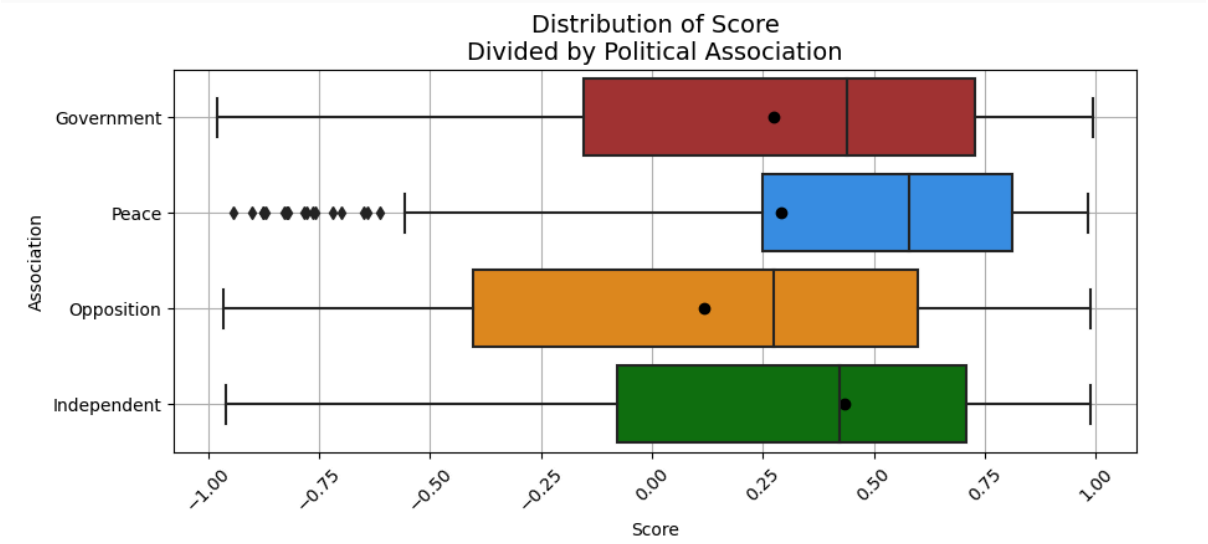
Here we can see that both women and men have tweets raging through the entire range of possible scores, though still showing a slight preference for tweets that can be classified as positive. In general, both distributions seem fairly close to one another, with a mean and a median only slightly more positive for women than for men, and a slightly wider range for the middle 50% on the male distribution. This would seem to go against common beliefs regarding men being significantly more aggressive than women.

Figure 6



By Political Association

Figure 7



In contrast to the relative cohesion when analyzed by gender, we see significantly more variation at the Political Association level. According to Figure 7, though the members belonging to the Peace Association seem to be the ones that have the higher proportion of their tweets classified as positive, it is the members belonging to the Independent Association who, on average tend to have more tweets that can be classified as positive.

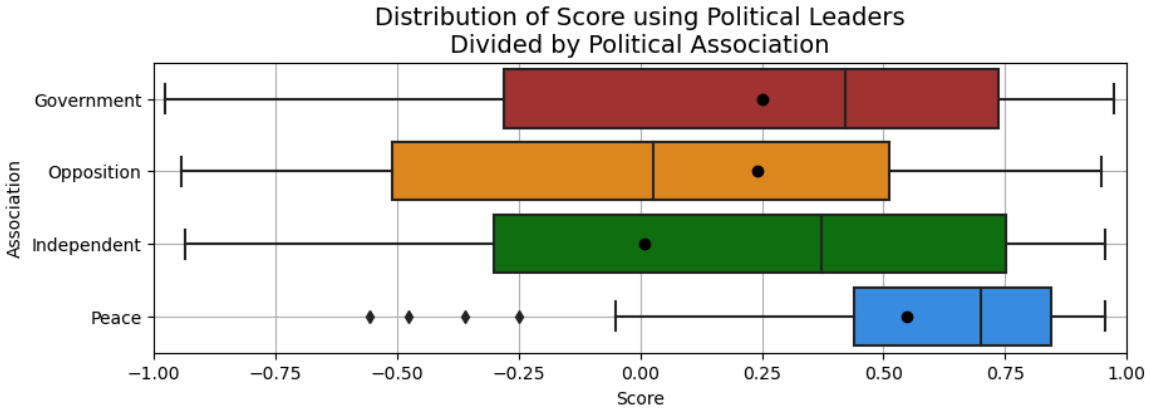
As for the differences between the scores for the tweets belonging to members grouped within the Opposition Association, as opposed to those of members grouped within

the Government Association, we see both a higher proportion of their overall tweets being classified as negative, as well as lower average scores.

It is, however, worth pointing out that, though considerable political turmoil was happening during this period, such as i) the end of a highly controversial presidency by Iván Duque, belonging to one of the parties of the current opposition, ii) the appointment of a new president, which not only is the first left-leaning politician to fill this place, but who was also involved within the guerrilla movement M-19, as well as other confounding factors, all parties have more than half of their tweets being classified as positive.

Use of Political Leaders

Figure 8



In this section, a closer look is taken at those tweets that were detected to contain mentions of the four most recent presidents[24]. In particular, Figure 8 helps us represent the distribution of the score, as well as the average score, for these types of tweets. Here we have that, excluding those associated with the Peace Association, the scores associated with these tweets expand through the entire possible range. Some key highlights from this graph include the fact that, true to their namesake, the Independent Association, despite seeming to feature mostly tweets with an associated positive score has an average score of close to zero.

[24] This refers to former Presidents (by term order) Alvaro Uribe Velez (perceived leader of the opposition), Juan Manuel Santos, Ivan Duque, and the current President, Gustavo Petro.

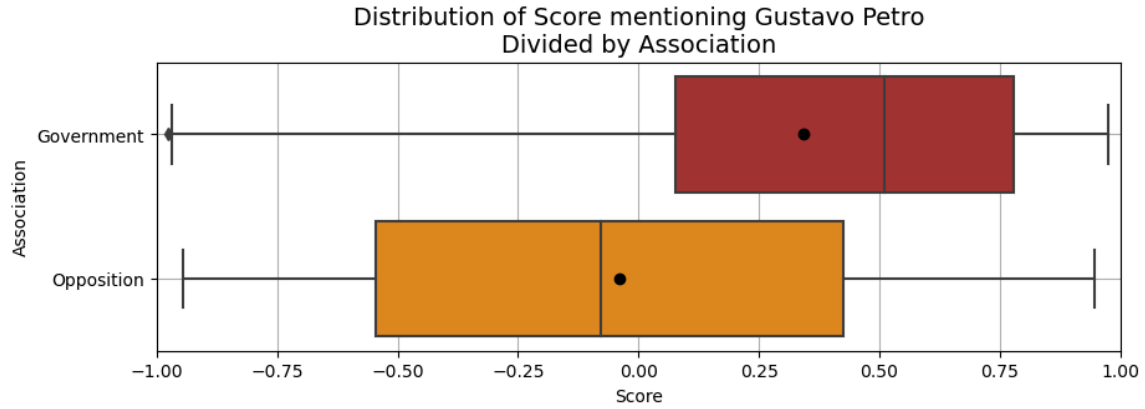
Also worth noting that, despite the Opposition Association featuring more negative scores than the Government Association, ultimately their average scores are very close. It is, of course, worth keeping in mind that being mentioned within an overall negative context does not necessarily imply that something negative is being said regarding that person, and the same should hold regarding being mentioned within an overall positive context.

Additionally, it is also relevant to mention that both the Peace Association, as well as the Independent Association, have a considerably lower number of tweets mentioning any of these political leaders, and even some missing values, meaning that any distribution we might see from them, will ultimately be lead by a small number of tweets, which given the scarcity, seem in turn seem to be quite different from the rest of their respective Associations, and hence, hardly fit to adequately represent them.

Because of this, the subsequent analysis will overlook these two associations, and instead concentrate on the differences between the Opposition Association and the Government Association, as both have a higher number of tweets compared to the other two, and a similar number of tweets amongst themselves.

Gustavo Petro: From Opposite Sides of the Political Spectrum

Figure 9



As expected, Figure 9 shows the Opposition Association featuring a higher number of tweets associated with negative scores, while the Government Association features a higher number of tweets associated with positive scores.

However, it is also worth noting that neither distribution seems particularly inclined towards either extreme, as shown by the fact that the average associated score for the Opposition Association is very close to being neutral, while the average associate score for the Government Association is still within the lower third of the positive distribution.

Further research might be necessary, however, to understand if this is the actual distribution, or if given that this is specifically filtering for their name, it is therefore ignoring cases in which members of the Opposition Association might use nicknames or even insults to refer to him, hence explaining the seeming lack of significant negative scores.

Alvaro Uribe[25]: From Opposite Sides of the Political Spectrum

Figure 10

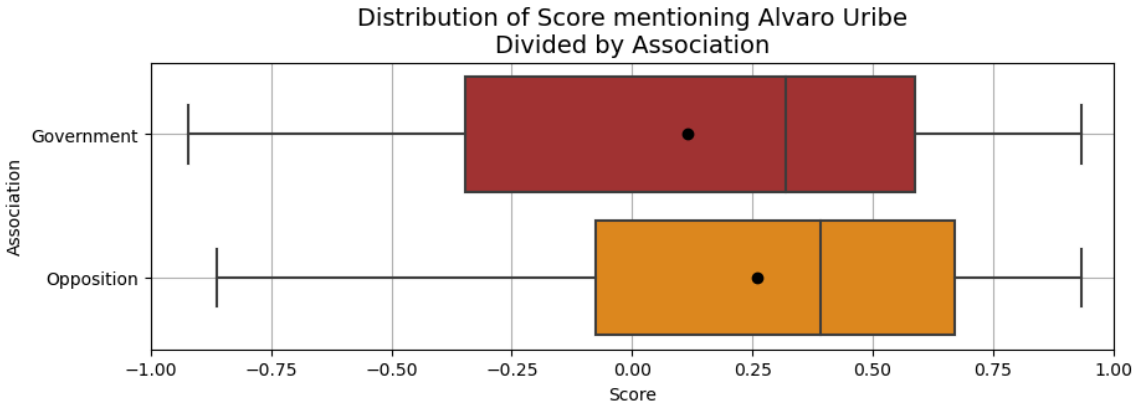


Figure 10 again, seems to stay within the expectations for the Opposition, as it shows how almost ¾ of the tweets from them have a positive score associated, as well as a positive average associated score. More unexpected, however, is the distribution for the Government

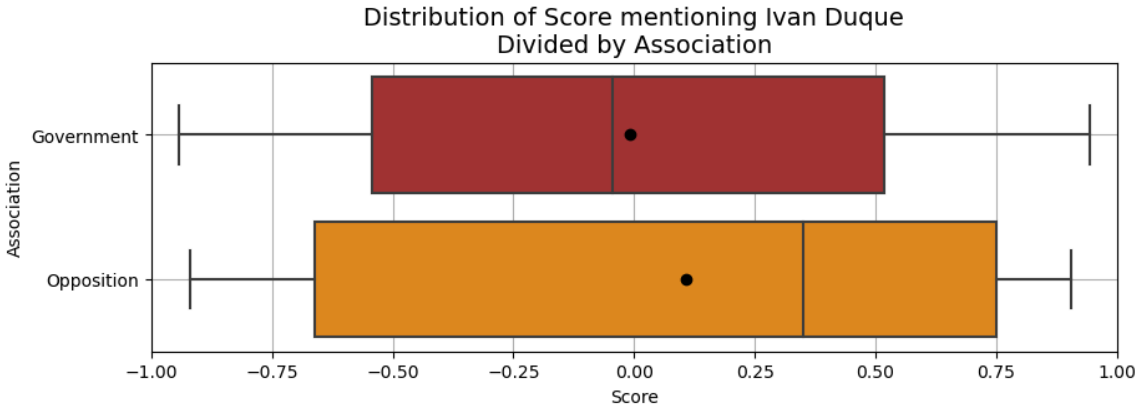
[25] First Presidential Term 2002 to 2006, Second Presidential Term from 2006 to 2010.

Association, not only do the associated scores cover the entire range, but more than half of the sample is positive, leading to an overall positive average score.

As is the case for the above result, further research should be conducted to determine whether this is the actual distribution, or if given that this is specifically filtering for their name, it is therefore ignoring cases in which members of the Government Association might use nicknames or even insults to refer to him, hence explaining the seeming lack of significant negative scores.

Iván Duque: From Opposite Sides of the Political Spectrum

Figure 11



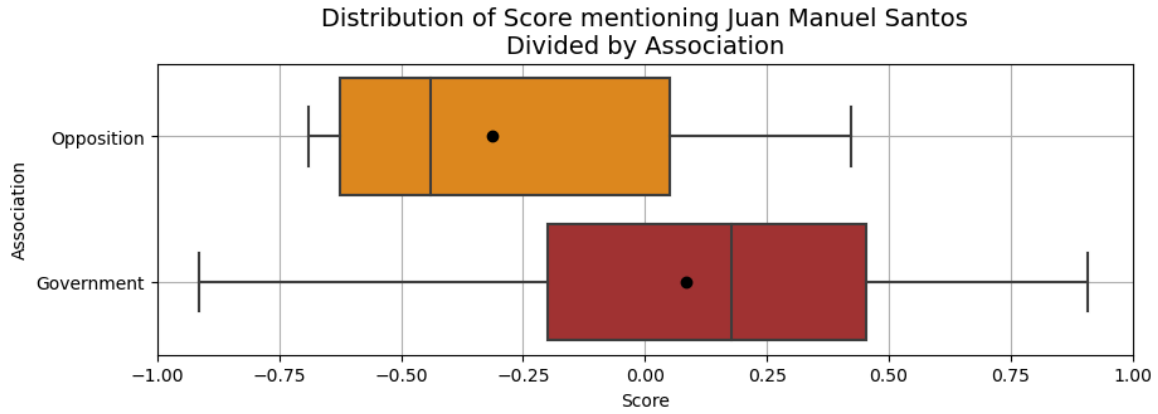
The figure of Ivan Duque, both within his party, as well as outside it, is a very complicated and controversial one. This is in large part due to his quite problematic presidential term, in which he was often ridiculed within social media for statements he made, as well as heavily critiqued inside and outside his party, over the decisions he took.

It is then no surprise that quite a large proportion of the score range is contained by the middle 50% of the data, both for the Opposition Association as well as for the Government Association. It is considered additionally telling that both the averages for the associated score for the Opposition and the Government seem to be close to one another, and

in turn, close to zero, though naturally the first is on the positive side of the spectrum, while the latter is on the negative one.

Juan Manuel: From Opposite Sides of the Political Spectrum

Figure 12



Lastly, there is the figure of Juan Manuel Santos, who had a quite complicated history during his two terms as President. After being the head of the Defense Ministry during the presidency of former President Alvaro Uribe, and being nominated by him to run as his successor, he decided near the end of this first term to break away from him by announcing he had been secretly negotiating what would later be known as the Peace Treaty of 2016 between the Colombian government and the guerrilla organization FARC-EP. These peace talks, were in turn, immediately condemned by former President Alvaro Uribe, and caused a political rift between the two figures.

Given this context, it is unsurprising that Figure 12 shows just a little under ¾ of the entire set of tweets from the opposition having a negative score associated with them, as well as the biggest average associated score in absolute terms, well into the negatives. More surprising though, is how the tweets within the Government Association are in turn mostly

contained within the positive range of the possible associated scores, as well as, very moderate, positive average associated score.

Public Response In terms of Score

The following plots and regressions seek to further add to the study of the score associated with each tweet, with the closest metric we have to evaluate viewer support, the number of likes. This analysis ultimately showed, after controlling for extreme values within the likes variable, as well as for possible inaccuracies within the score variable, a negative relationship between the score associated with each tweet, and the number of likes, which would suggest a preference for negative over positive content.

Additionally, these results held, both in the case of analyzing the entire range for the associated score, as well as when analyzing the negative and the positive sub-sections. Ultimately, though not every coefficient was classified as statistically significant, all those that were, agreed with this particular trend.

By Gender

Figure 13

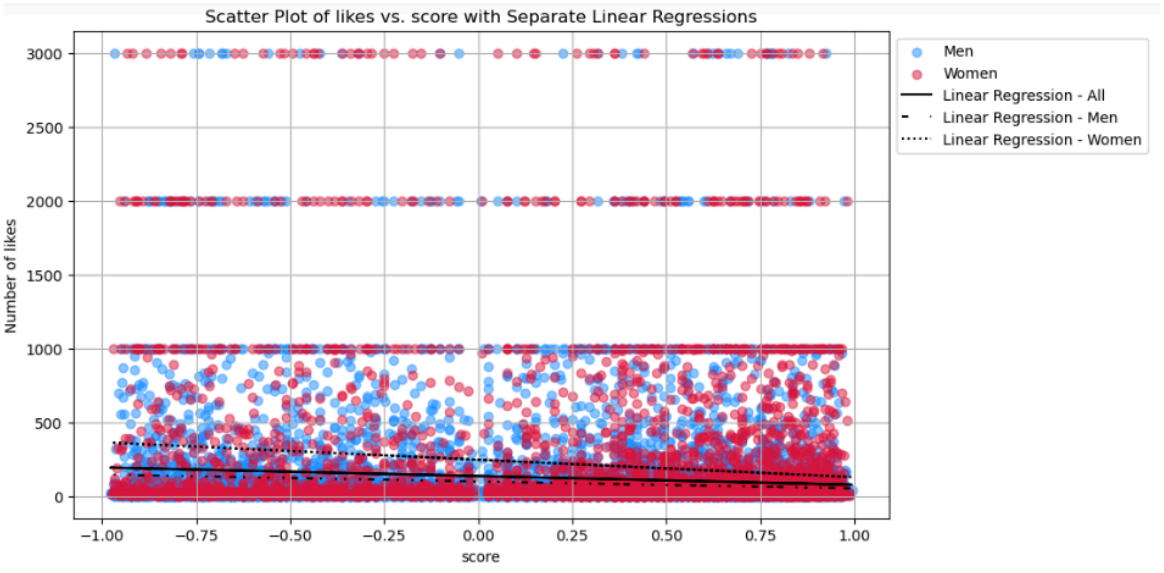
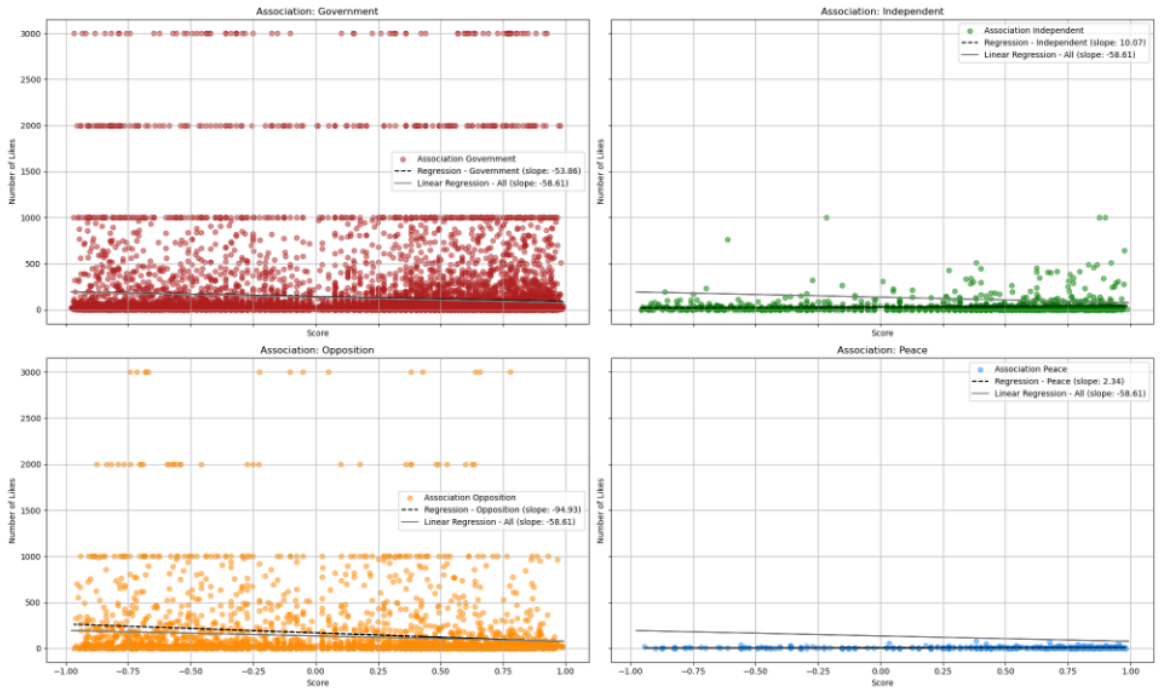


Figure 13 shows a scatterplot, mapping each point according to its associated score and the number of likes it received, additionally, points have been colored either pink for women or blue for men. Finally, a linear regression was run with the overall data, as well as with the data segregated by gender.

For the General Regression, the **General Coefficient was -58.61**, with an **intercept of 135.98** and a **p-value of 2.40e-50**. This means that for our overall sample, there was a negative relation between the Score and the number of likes, which was statistically significant. Additionally, for women, the **Coefficient was -203.74**, considerably higher than the general one, with an **intercept of 217.67** and a of **p-value of 0.003**. Finally, the results for men show a **Coefficient of -49.21**, with an **intercept of 99.70** and a **p-value of 0.017**.

By Political Association

Figure 14



Similar to Figure 14, Figure 15 shows the relationship between the number of likes and the associated score of each tweet, separated by the political association of the author responsible for each tweet. For the Political Associations, we find that:

- ➔ The Coefficient for Government is: Slope: -53.86, Intercept: 142.59, p-value: 0.0000
- ➔ The Coefficient for Independent is: Slope: 10.07, Intercept: 29.25, p-value: 0.0107
- ➔ The Coefficient for Opposition is: Slope: -94.93, Intercept: 169.01, p-value: 0.0000
- ➔ The Coefficient for Peace is: Slope: 2.34, Intercept: 7.84, p-value: 0.0138

In general, these results seemed to indicate two conflicting effects, for parties contained within either the Government Association or the Opposition Association, there is a negative relation between the number of likes and the associated score for a particular tweet, meanwhile, parties contained within the Peace Association or Independent Association there is a positive effect between the number of likes and the associated score for a particular tweet.

Conclusions

Overall, there seem to be 3 main takeaways from these results. Starting with the fact that, though the gender representation gap has shown improvement, much is still necessary to achieve true equality. Also worth noting is the significant shift traditional parties have made, and though it is ultimately unclear the solidity of said shift, the simple existence of it is already much more than many imagined likely.

On the Sentiment Analysis, political speech seemed significantly more positive than expected, and the handling of potentially controversial subjects such as past political leaders, while still being influenced by the underlying political associations, seemed to be considerably more moderated than would otherwise be expected.

Finally, in terms of public interactions with tweets, the fact that the number of likes a particular tweet received is negatively related to the associated score said tweet received,

could be a reflection of the inner turmoil that the public feels, which is, to a lesser extent, played out within public arenas such as social media. Additionally, given the lack of a random variation, as well as the fact that the Independent and Peace Coalitions have significantly smaller samples compared to the Government and Opposition Associations, it is hard to try and draw any significant conclusions from the differences in the effects. Given this, a potentially fruitful continuation of this research could include collecting the entire sample for the first three months, or even a longer period, and diving deeper into this relation.

Ultimately, it is clear that more research and attention into this topic is required to get a better grasp and understanding of how politicians and civilians interact within the social media sphere. Subsequent research could also consider further delving into the vocabulary utilized across political party lines, utilizing tools such as topic modeling.

Acknowledgments

I would like to utilize this space to thank all my loved ones, whose support was paramount for the completion of this thesis. I would also like to thank my professors at the Hertie School, at *Universidad del Rosario*, and beyond, for giving me both the tools that have ultimately led me to this position, but more importantly, the drive to do something with them.

On a more formal note, I would like to thank Universidad de los Andes for making available their webpage *Congreso Visible*, not just because of the public service it provides, but because without it, this endeavor would have been entirely unfeasible.

References

Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, *10*(2), 361–374.

Baccouri, N. (2020). *Deep_translator* (1.5.4) [Computer software]. https://deep-translator.readthedocs.io/en/latest/index.html

Barreto-Galeano, M. I., Medina-Arboleda, I. F., Zambrano-Hernández, S., Sabucedo-Cameselle, J.-M., Blanco-Abarca, A., & Maurice Lair, É. (2021). Rhetoric, Political Ideology and the Peace Process in Colombia: A Twitter® Analysis. *Studies in Conflict & Terrorism*, *44*(11), 920–937. https://doi.org/10.1080/1057610X.2019.1615256

Bonilla, J. M. H. (2022, September 8). *Petro sella una sólida coalición en el Congreso con el apoyo de los partidos tradicionales*. El País América Colombia. https://elpais.com/america-colombia/2022-09-08/petro-sella-una-solida-coalicion-en-el-congreso-con-el-apoyo-de-los-partidos-tradicionales.html

Botero, M. C. (2022, July 20). El difícil viaje hacia la paridad de género en el Congreso de Colombia. *Pesquisa Javeriana*. https://www.javeriana.edu.co/pesquisa/el-dificil-viaje-hacia-la-paridad-de-genero-en-el-congreso-de-colombia/

Buccoliero, L., Bellio, E., Crestini, G., & Arkoudas, A. (2020). Twitter and politics: Evidence from the US presidential elections 2016. *Journal of Marketing Communications*, *26*(1), 88–114. https://doi.org/10.1080/13527266.2018.1504228

Cámara De Representantes. (n.d.). Retrieved April 27, 2024, from https://www.camara.gov.co

Cámara De Representantes: Mapa Electoral. (n.d.). Retrieved April 29, 2024, from https://www.camara.gov.co

Ceron, A. (2017). Social Media, Political Institutions and the Principal-Agent Dilemma. In

Social Media and Political Accountability: Bridging the Gap between Citizens and Politicians (pp. 1–11). Springer.

Colombia Renaciente. (2022). *Colombia Renaciente se declara Partido de Gobierno ante el Consejo Nacional Electoral.* https://www.partidocolombiarenaciente.co/web/wp-content/uploads/2022/12/COM_DECLARACION-PARTIDO-GOBIERNO_Julio032022-1.pdf

Conoce a todos los congresistas 2022—2026 de Colombia. (n.d.). Congreso Visible. Retrieved April 27, 2024, from https://congresovisible.uniandes.edu.co/congresistas

Gallego, J., Martínez, J. D., Munger, K., & Vásquez-Cortés, M. (2019). Tweeting for peace: Experimental evidence from the 2016 Colombian Plebiscite. *Electoral Studies*, *62*, 102072.

Galvis, J. E. L., Manuela. (2022, July 19). *Así es el Congreso 2022-2026*. La Silla Vacía. http://www.lasillavacia.com/silla-nacional/asi-es-el-congreso-que-se-instala-manana/

Giachanou, A., & Crestani, F. (2017). Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Computing Surveys*, *49*(2), 1–41. https://doi.org/10.1145/2938640

Haman, M., Školník, M., & Čopík, J. (2022). Colombian political leaders on Twitter during the Covid-19 pandemic. *Latin American Policy*, *13*(1), 104–121. https://doi.org/10.1111/lamp.12249

Herre, B., Rodés-Guirao, L., Ortiz-Ospina, E., & Roser, M. (2024). Democracy. *Our World in Data*. https://ourworldindata.org/democracy

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://ojs.aaai.org/index.php/icwsm/article/view/14550

Infobae. (2022, September 21). *Qué partidos políticos están en el Gobierno de Petro y cuáles en la oposición: CNE dio su fallo definitivo*. infobae. https://www.infobae.com/america/colombia/2022/09/21/que-partidos-politicos-estan-en-el-gobierno-de-petro-y-cuales-en-la-oposicion-cne-dio-su-fallo-definitivo/

La Silla Vacía. (2024, March 22). *Quién es Juan Sebastián Gómez Gonzáles*. La Silla Vacía. http://www.lasillavacia.com/quien-es-quien/quien-es-juan-sebastian-gomez-gonzales/

Ley 581 de 2000 (May 31, 2000). https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=5367

Ley 1909 de 2018 (July 9, 2018). https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=87302

Merino, A. D., & López Barajas, P. (n.d.). *Mujeres y Hombres: Brechas de Género en Colombia*. https://www.dane.gov.co/files/investigaciones/genero/publicaciones/mujeres-y-hombre-brechas-de-genero-colombia-informe.pdf

Muñoz, M. I. M. (2022, March 15). *Curules para la paz: Estas son las 16 personas que van a llegar a ocuparlas*. El Tiempo. https://www.eltiempo.com/elecciones-2022/congreso/curules-de-paz-estos-son-los-representantes-electos-658393

Redacción Política. (2022, July 15). *Gobierno, independencia u oposición: La postura de los partidos frente a Petro* [Text]. ELESPECTADOR.COM. https://www.elespectador.com/politica/gobierno-independencia-u-oposicion-la-postura-de-los-partidos-frente-al-gobierno-petro/

Rodríguez, D. P. (2023, February 26). Cuatro de cada diez colombianos gastan su tiempo libre en redes sociales. *Diario La República*. https://www.larepublica.co/ocio/en-enero-casi-75-de-los-colombianos-usaron-las-redes-sociales-3554927

Schmitter, P. C. (2007). Political accountability in 'real-existing'democracies: Meaning and mechanisms. *Florence: European University Institute. Recuperado Em*, *8*.

Statement of Authorship

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated. I confirm that the digital copy of the master thesis that I submitted on April 29th, 2024 is identical to the printed version I submitted to the Examination Office on April 20th, 2024.

DATE: April 29th, 2024

NAME: María José de Jesús Lee Ocampo

SIGNATURE: María José Lee