

**Exploración de la percepción pública en X sobre el sistema tributario, aduanero y cambiario de Colombia, a partir del análisis de 1.000 publicaciones relacionadas con la DIAN, Ministerio de Hacienda y Banco de la República (octubre de 2025)**

Hans Mateo Morantes Maldonado

Asesor

Jorge Luis Quintero López

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

## Resumen

En la era digital, las redes sociales se han consolidado como un canal esencial de comunicación entre el Estado y la ciudadanía. No obstante, muchas entidades públicas en Colombia carecen de mecanismos estructurados para medir la percepción ciudadana frente a temas de su gestión, lo que dificulta evaluar la efectividad de sus estrategias comunicativas y limita la capacidad institucional de respuesta ante opiniones negativas o críticas recurrentes.

El presente proyecto tiene como propósito explorar la percepción pública en la red social X (antes Twitter) frente a temas relacionados con el sistema tributario, aduanero y cambiario de Colombia, mediante el análisis de 1.000 publicaciones asociadas a las entidades Dirección de Impuestos y Aduanas Nacionales (DIAN), Ministerio de Hacienda y Crédito Público y Banco de la República, recolectadas durante el mes de octubre de 2025 a través de la API oficial de X.

Para el procesamiento y análisis de los datos se emplearon herramientas de Python, incluyendo las librerías pandas, matplotlib y pysentimiento, con el fin de realizar una clasificación automática del sentimiento (positivo, negativo o neutro) y visualizar las tendencias generales en las reacciones ciudadanas. Además, se aplicó una validación manual del 15 % del corpus para evaluar la precisión del modelo y fortalecer la confiabilidad de los resultados obtenidos.

Los hallazgos derivados de esta investigación buscan apoyar la mejora de la comunicación institucional, ofreciendo una base empírica para adaptar mensajes según la percepción pública y fortalecer la relación entre el Estado y la ciudadanía en el entorno digital.

**Palabras clave:** X (antes Twitter), percepción ciudadana, comunicación institucional, sistema tributario, sistema aduanero, sistema cambiario, análisis de sentimiento, procesamiento de lenguaje natural (PLN), DIAN, Ministerio de Hacienda, Banco de la República

## Abstract

In the digital era, social media has become an essential communication channel between the State and citizens. However, many public entities in Colombia lack structured mechanisms to measure public perception regarding issues related to their management, which makes it difficult to evaluate the effectiveness of their communication strategies and limits the institutional capacity to respond to negative opinions or recurring criticism.

The purpose of this project is to explore public perception on the social network X (formerly Twitter) regarding topics related to Colombia's tax, customs, and foreign exchange systems, through the analysis of 1,000 posts associated with the entities Dirección de Impuestos y Aduanas Nacionales (DIAN), Ministerio de Hacienda y Crédito Público, and Banco de la República, collected during October 2025 using X's official API.

For data processing and analysis, Python tools were used, including the libraries *pandas*, *matplotlib*, and *pysentimiento*, in order to perform automatic sentiment classification (positive, negative, or neutral) and visualize general trends in citizen reactions. In addition, a manual validation of 15% of the corpus was conducted to evaluate the model's accuracy and strengthen the reliability of the obtained results.

The findings derived from this research aim to support improvements in institutional communication by providing an empirical basis to adapt messages according to public perception and strengthen the relationship between the State and citizens in the digital environment.

**Keywords:** X (formerly Twitter), public perception, institutional communication, tax system, customs system, foreign exchange system, sentiment analysis, natural language processing (NLP), DIAN, Ministry of Finance, Banco de la República.

## Tabla de Contenido

Justificación .....	6
Objetivos.....	8
Objetivo General .....	8
Objetivos Específicos.....	8
Marcos de Referencia .....	9
Marco Conceptual .....	9
Marco Teórico.....	10
Metodología .....	12
Resultados Obtenidos.....	14
Conclusiones .....	29
Referencias Bibliográficas .....	31

## Lista de Figuras

<b>Figura 1</b> <i>Script para Extraer Publicaciones de la DIAN desde la API de X</i> .....	14
<b>Figura 2</b> <i>Script para la Limpieza de los Datos</i> .....	15
<b>Figura 3</b> <i>Script para la Limpieza de los Datos</i> .....	16
<b>Figura 4</b> <i>Distribución General de Sentimientos</i> .....	17
<b>Figura 5</b> <i>Distribución de Sentimientos por Entidad</i> .....	18
<b>Figura 6</b> <i>Engagement Promedio por Sentimiento</i> .....	19
<b>Figura 7</b> <i>Engagement Promedio por Sentimiento en la DIAN</i> .....	20
<b>Figura 8</b> <i>Engagement Promedio por Sentimiento en el Banco de la República</i> .....	21
<b>Figura 9</b> <i>Engagement Promedio por Sentimiento en el Ministerio de Hacienda</i> .....	22
<b>Figura 10</b> <i>Top 20 Usuarios por Número de Tweets</i> .....	23
<b>Figura 11</b> <i>Pareto del Engagement Total por Tweet</i> .....	24
<b>Figura 12</b> <i>Proporción de Sentimientos en el Top 35 de Tweets por Engagement</i> .....	25
<b>Figura 13</b> <i>Nube de Palabras para el Sentimiento Positivo</i> .....	26
<b>Figura 14</b> <i>Nube de Palabras para el Sentimiento Negativo</i> .....	27
<b>Figura 15</b> <i>Nube de Palabras para el Sentimiento Negativo</i> .....	28

## Justificación

Este estudio es relevante porque permitirá comprender cómo percibe la población colombiana el sistema tributario, aduanero y cambiario, a partir del análisis de publicaciones realizadas por los usuarios en la red social X (antes Twitter). Estas publicaciones reflejan opiniones y valoraciones espontáneas sobre el funcionamiento de las principales entidades que integran este sistema —la DIAN, el Ministerio de Hacienda y Crédito Público y el Banco de la República—, lo que convierte a la plataforma en un espacio valioso para el análisis de la percepción pública.

En Colombia, la mayoría de las instituciones estatales no dispone de mecanismos automatizados ni estructurados que permitan medir de forma continua las reacciones ciudadanas frente a temas relacionados con su gestión. Esta ausencia de herramientas limita la capacidad institucional para anticipar percepciones negativas, identificar focos de descontento o reconocer aspectos valorados por la población.

El proyecto propone aplicar técnicas de análisis de sentimiento y procesamiento de lenguaje natural (PLN) en Python, mediante librerías como pandas, matplotlib y pysentimiento, sobre un corpus de 1.000 publicaciones recolectadas durante octubre de 2025 a través de la API oficial de X. Este enfoque permitirá clasificar las reacciones ciudadanas (positivas, negativas o neutras) y visualizar las tendencias generales en torno a los temas tributarios, aduaneros y cambiarios.

Los resultados aportarán una visión empírica y cuantificable de la percepción pública hacia las entidades fiscales y financieras del país, sirviendo como insumo para fortalecer la comunicación institucional, promover estrategias de interacción más cercanas y fomentar una relación Estado-ciudadanía basada en la escucha activa y el análisis de datos. En el ámbito

académico, el estudio contribuye al desarrollo de metodologías aplicadas de análisis de opinión pública digital en español, incorporando herramientas de procesamiento de lenguaje natural al análisis gubernamental.

## Objetivos

### Objetivo General

Explorar la percepción pública frente a las publicaciones en X relacionadas con el sistema tributario, aduanero y cambiario mediante técnicas de análisis de sentimiento con Python (pandas, matplotlib y pysentimiento), para describir patrones y características generales de las reacciones ciudadanas.

### Objetivos Específicos

Recolectar 1000 publicaciones de X relacionadas con el sistema tributario, aduanero y cambiario, así como aquellas realizadas por usuarios que mencionan o interactúan con las principales entidades del sistema (DIAN, Ministerio de Hacienda y Banco de la República), utilizando la API de X a través de la librería *requests* en Python, con el fin de conformar el corpus de análisis.

Depurar y estructurar el corpus de 1000 tweets recolectados, mediante procesos de limpieza, normalización y consolidación en Python (usando pandas y expresiones regulares), garantizando la coherencia de las variables y la calidad de los datos para el análisis de sentimiento.

Aplicar técnicas de procesamiento de lenguaje natural y análisis de sentimiento en español, utilizando la librería *pysentimiento* en Python, con el fin de clasificar las publicaciones recolectadas según la reacción ciudadana (positiva, negativa o neutra).

Explorar los resultados obtenidos del análisis de sentimiento y de la frecuencia de términos en los tweets recolectados, con el propósito de reconocer patrones generales en la percepción ciudadana.

## Marcos de Referencia

### Marco Conceptual

Este trabajo se centra en el análisis de la percepción ciudadana frente a las publicaciones relacionadas con las principales entidades del sistema tributario, aduanero y cambiario en Colombia en Twitter (X) a partir de un dataset de 1000 tweets, mediante técnicas de análisis de sentimiento. Se define este fenómeno desde varias perspectivas:

1. Percepción ciudadana: es la manera en la que los ciudadanos interpretan, valoran y reaccionan ante los mensajes institucionales. Puede incluir actitudes explícitas o implícitas, influenciadas por contexto social, emocional y cultural (Vizmanos et al., 2023; Figueroa Perdomo, 2020).
2. Comunicación gubernamental digital: conjunto de prácticas y estrategias utilizadas por entidades estatales para interactuar con la ciudadanía en línea, incluyendo contenido, lenguaje, tiempo de publicación y formas de respuesta (Ruiz Soto, 2023; Arango Espinal et al., 2025).
3. Análisis de sentimiento: técnica de PLN que determina la polaridad emocional del texto (positivo, negativo, neutral), útil para cuantificar reacciones a publicaciones. Se suelen usar herramientas como TextBlob y VADER (en Python) así como métodos lexicon-based o basados en aprendizaje automático (Mat Razali et al., 2021; Moreno-Sandoval et al., 2024). Su aplicación en Twitter ha sido consistente: se detecta mayor viralidad de mensajes negativos en contextos políticos o institucionales (Antypas et al., 2022)
4. Minería de opinión: va más allá de la polaridad, capturando emociones finas, temas recurrentes y valoraciones subyacentes (Suhaimin et al., 2023; Antypas et al., 2022). En

contextos institucionales, permite detectar qué aspectos del mensaje (imagen, contenido, tono) generan mayor impacto o rechazo.

5. Modelos de análisis complementarios: se usan herramientas como topic modeling (LDA) que ayudan a visualizar temas emergentes en los tuits o comentarios asociados (Ghassan AbuRaed et al., 2024; Conti et al., 2023). Así como modelos avanzados (BERT-CNN, Transformers) mejoran la precisión en análisis de texto breve, relevante en Twitter (Joloudari et al., 2022; Iguarán Fernández et al., 2024)

6. Ética y validación: la adaptación local (español colombiano) es esencial: los modelos traducidos directamente pueden inducir sesgos o malas interpretaciones (Venkit & Wilson, 2021; Moreno-Sandoval et al., 2024). Requiere evaluación contextualizada para asegurar representatividad y confiabilidad del análisis desarrollado.

### **Marco Teórico**

Comunicación Digital e Influencia Institucional: Figueroa Perdomo (2020) y Arango Espinal et al. (2025) explican cómo las redes institucionales en redes sociales pueden legitimar o debilitar al Estado, particularmente en temas como impuestos o presupuesto, dependiendo de la percepción que se genere. Además, Quinchía-Saavedra et al. (2025) encuentran que mensajes negativos en Twitter durante procesos como el posconflicto generan más retuits, lo que evidencia la capacidad viral de emociones negativas.

Técnicas de PLN y herramientas: Mat Razali et al. (2021) y Suhaimin et al. (2023) validan técnicas léxicas (TextBlob, VADER) para contextos institucionales; Moreno-Sandoval et al. (2024) crean recursos semánticos específicos para español colombiano, mejorando de manera importante la precisión. Joloudari et al. (2022) muestran que modelos híbridos (BERT-CNN) que superan enfoques tradicionales para textos breves, lo cual es clave en análisis de tuits. También

existen Architectural frameworks como TweetNLP que ofrecen modelos optimizados para redes sociales, incluyendo detección de sentimientos y discurso de odio

Dinámica Emocional y Viralidad: Antypas, Preece y Camacho-Collados (2022) demostraron la mayor viralidad de contenido negativo, lo cual es crucial para entender cómo ciertos temas institucionales se amplifican. Iguarán Fernández et al. (2024) proponen detección de emociones finas durante eventos electorales, identificando escala emocional compleja

## Metodología

El desarrollo del proyecto se fundamentó en una metodología de carácter aplicado, basada en técnicas de ciencia de datos y procesamiento de lenguaje natural (PLN). El enfoque combinó procedimientos automatizados de análisis con etapas de validación manual, asegurando tanto la eficiencia técnica como la rigurosidad analítica.

Se extrajeron 1.000 publicaciones realizadas durante octubre de 2025 en la red social X (antes Twitter), relacionadas con el sistema tributario, aduanero y cambiario de Colombia. Para ello, se utilizó la API oficial de X junto con la librería requests de Python, aplicando un conjunto de palabras clave y etiquetas asociadas a las principales entidades del sistema: la Dirección de Impuestos y Aduanas Nacionales (DIAN), el Ministerio de Hacienda y Crédito Público y el Banco de la República.

Los datos recolectados fueron depurados y estandarizados en Python mediante el uso de pandas y expresiones regulares. Este proceso incluyó la eliminación de duplicados, la normalización del texto, la corrección de codificación, la remoción de URLs, menciones y hashtags, además de la tokenización del contenido textual para su posterior análisis.

Posteriormente, se aplicaron técnicas de análisis de sentimiento a través de las librerías pysentimiento, entrenadas sobre modelos de PLN en español. Cada publicación fue clasificada en tres categorías de sentimiento: positivo, negativo o neutro. Los resultados obtenidos fueron complementados con visualizaciones exploratorias y distribuciones de sentimiento por entidad, así como con análisis de frecuencia de términos y nubes de palabras.

Para garantizar la confiabilidad de los resultados, se llevó a cabo una validación manual del 15 % del corpus, verificando la concordancia entre la clasificación automática y la interpretación humana. A partir de esta revisión, se sistematizaron los hallazgos y se

identificaron patrones generales en la percepción ciudadana, así como diferencias en las reacciones hacia cada entidad o tema analizado.

Finalmente, el flujo completo del análisis se implementó en Python utilizando las librerías pandas, matplotlib, wordcloud y pysentimiento. Los resultados fueron exportados a archivos Excel (.xlsx) e ilustrados mediante gráficos comparativos, series temporales y nubes de palabras, lo que permitió una interpretación visual y estructurada de la percepción ciudadana frente al sistema tributario, aduanero y cambiario colombiano.



Código de Python que permita realizar la limpieza de los datos recolectados de X y un archivo CSV o XLSX en el cual se encuentren los datos limpios

## Figura 2

### Script para la Limpieza de los Datos

```
import pandas as pd
import re
from pathlib import Path
from openpyxl import load_workbook
from openpyxl.worksheet.table import Table, TableStyleInfo

# 1. Ruta a La carpeta con archivos CSV
carpeta_csv = Path("C:/Users/User/Desktop/Codigos especialización/Proyecto de grado/Tweets descargados")

# 2. Acumulador de todos Los datos
df_total = pd.DataFrame()

# 3. Recorrer todas Los archivos CSV
for archivo_csv in carpeta_csv.glob("*.csv"):
    try:
        df = pd.read_csv(archivo_csv, quoting=1, encoding="utf-8-sig", on_bad_lines='skip')

        # Convertir a datetime y quitar timezone
        df["Fecha"] = pd.to_datetime(df["Fecha"], errors="coerce")
        df = df.dropna(subset=["Fecha"])
        df["Fecha"] = df["Fecha"].dt.tz_localize(None)

        # Eliminar filas sin texto
        df = df.dropna(subset=["Tweet"])

        # Agregar al total (sin columnas extra)
        df_total = pd.concat([df_total, df], ignore_index=True)

    except Exception as e:
        print(f" Error procesando {archivo_csv.name}: {e}")

# 4. Guardar como Excel
nombre_archivo = "tweets_dian_limpios.xlsx"
df_total.to_excel(nombre_archivo, index=False, sheet_name="Tweets")

# 5. Convertir hoja en tabla de Excel
wb = load_workbook(nombre_archivo)
ws = wb["Tweets"]

# Calcular rango de La tabla
ultima_columna = chr(65 + len(df_total.columns) - 1)
ultima_fila = str(len(df_total) + 1)
rango = f"A1:{ultima_columna}{ultima_fila}"

tabla = Table(displayName="TweetsDIAN", ref=rango)
estilo = TableStyleInfo(name="TableStyleMedium9", showFirstColumn=False,
                        showLastColumn=False, showRowStripes=True, showColumnStripes=False)
tabla.tableStyleInfo = estilo

ws.add_table(tabla)
wb.save(nombre_archivo)

# 6. Mostrar resultado
# =====
print(df_total.head())
print(f" Archivo Excel generado en formato de tabla: {nombre_archivo}")
print(f" Total de tweets procesados: {len(df_total)}")
```

*Nota.* La limpieza garantiza que los datos quedan en el formato indicado para realizar el análisis de sentimiento.

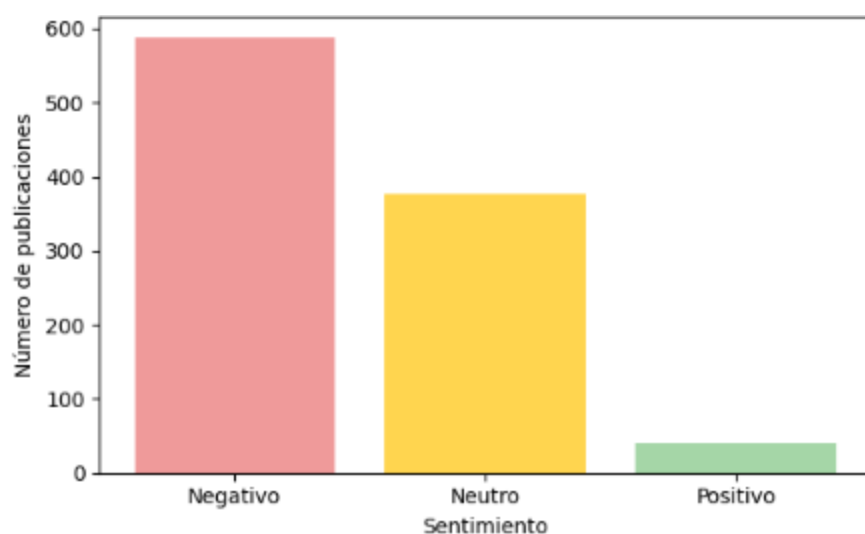


Código de python que permita realizar la debida exploración de los datos y presente gráficamente los insights más relevantes.

En la figura 4 se aprecia una prevalencia clara de opiniones negativas (587; 58,5%), seguidas por mensajes neutros (376; 37,5%) y una fracción menor positiva (40; 4,0%). Esta asimetría sugiere que, durante el periodo observado, las conversaciones en X sobre el sistema tributario, aduanero y cambiario se caracterizaron por críticas, inquietudes o insatisfacción más que por valoración favorable.

#### **Figura 4**

##### *Distribución General de Sentimientos*

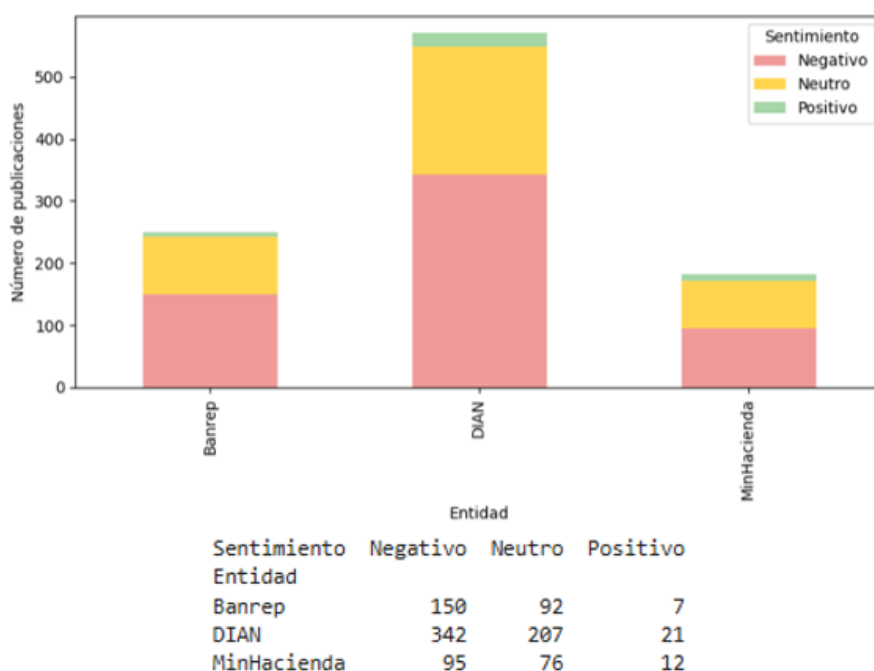


Esta estructura de polaridad tiene dos implicaciones: la primera es que el clima conversacional de referencia es predominantemente reactivo ante medidas y actores del ámbito fiscal y segundo que la baja proporción de positividad limita la existencia de narrativas pro-institucionales o de “casos de éxito” que logren tracción orgánica.

En la figura 5 se aprecia que, al segmentar por entidad, se mantiene la pauta de negatividad, pero en proporciones distintas

**Figura 5**

*Distribución de Sentimientos por Entidad*



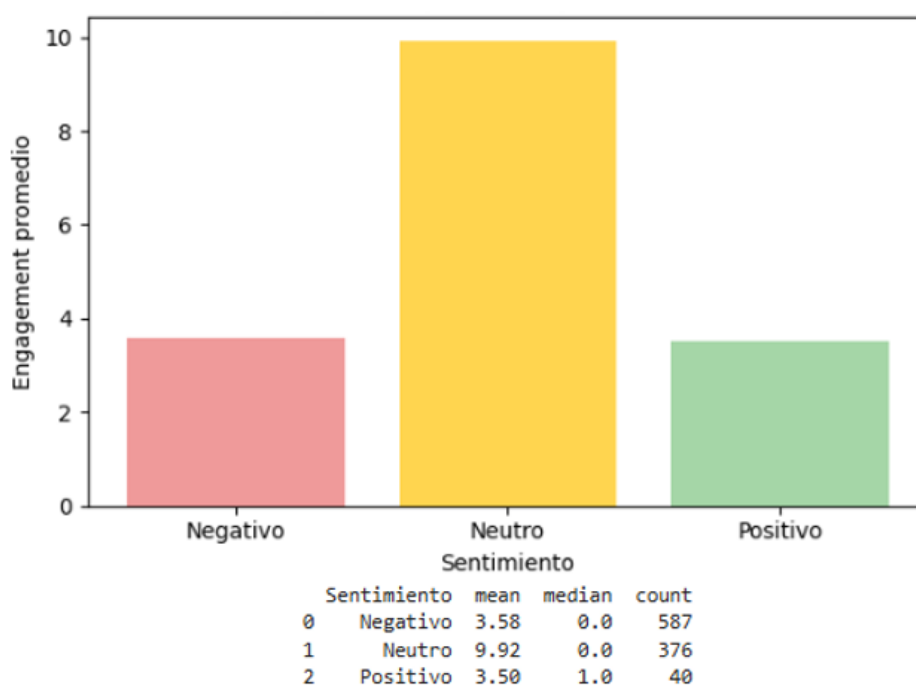
*Nota.* El gráfico muestra la proporción de publicaciones negativas, neutras y positivas asociadas a cada entidad analizada.

La DIAN concentra el mayor volumen de conversación y, en consecuencia, también la mayor cantidad de menciones negativas. El Banco de la Republica y el Ministerio de Hacienda conservan perfiles cualitativamente similares (predominio negativo), pero con volúmenes menores. En términos de gestión, esto indica que las estrategias de escucha y respuesta deberían priorizar a DIAN por escala (potencial reputacional) y a Banrep por la sensibilidad macroeconómica de sus temas (monetaria, inflación).

En la figura 6 se evidencia que el engagement promedio (likes + RTs + respuestas) es mayor para publicaciones neutras (mean≈9,92), en comparación con negativas (mean≈3,58) y positivas (mean≈3,50). Aunque el promedio favorece a lo neutro, la mediana cero en dos categorías evidencia una distribución altamente sesgada: pocos contenidos concentran la interacción, mientras que la mayoría recibe poca o ninguna.

**Figura 6**

*Engagement Promedio por Sentimiento*

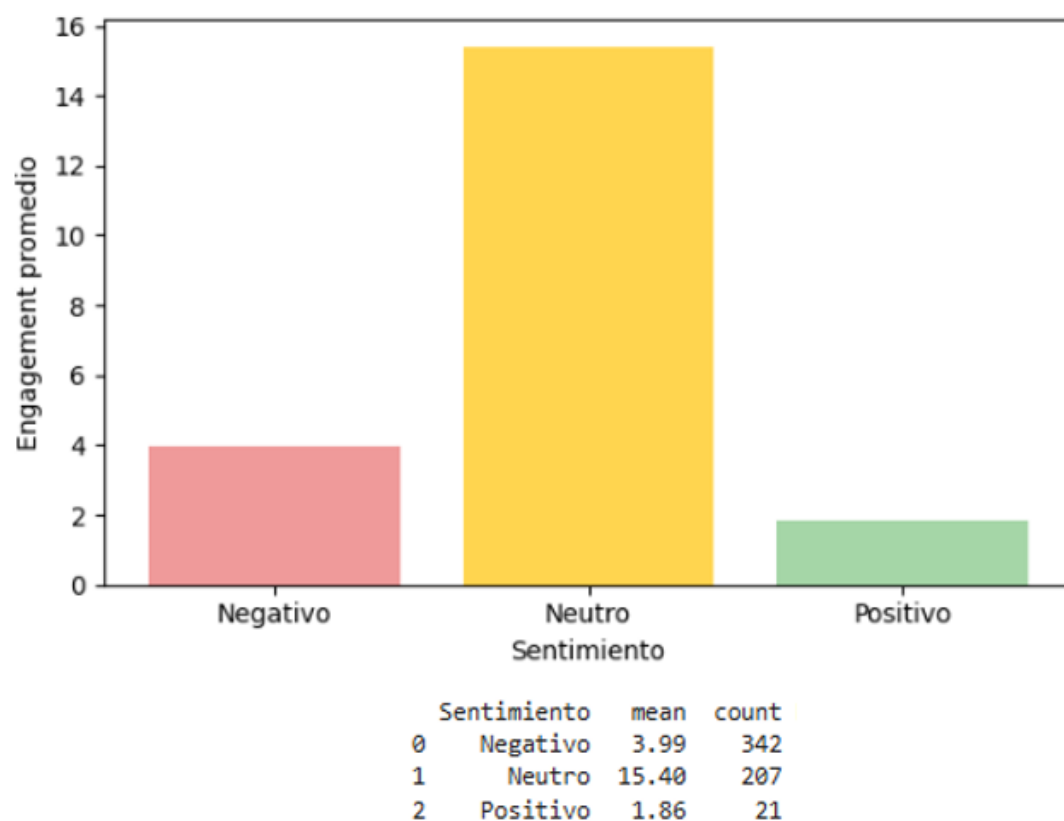


Al identificar estas publicaciones que generan más *engagement*, las entidades pueden seleccionar estratégicamente cuáles responder, enfocándose en aquellas que tienen mayor capacidad de influir en la conversación pública y, por tanto, de mejorar la percepción ciudadana mediante intervenciones oportunas y visibles.

En engagement por sentimiento de cada entidad se puede decir que el patrón se matiza por actor institucional. En la figura 7 se aprecia que, para el caso particular de la DIAN, las publicaciones con sentimiento neutro son las que más engagement tienen, lo cual concuerda con el engagement promedio por sentimiento, sin embargo, el engagement promedio en publicaciones negativas es considerablemente mayor al engagement de las publicaciones con sentimiento positivo

### Figura 7

*Engagement Promedio por Sentimiento en la DIAN*

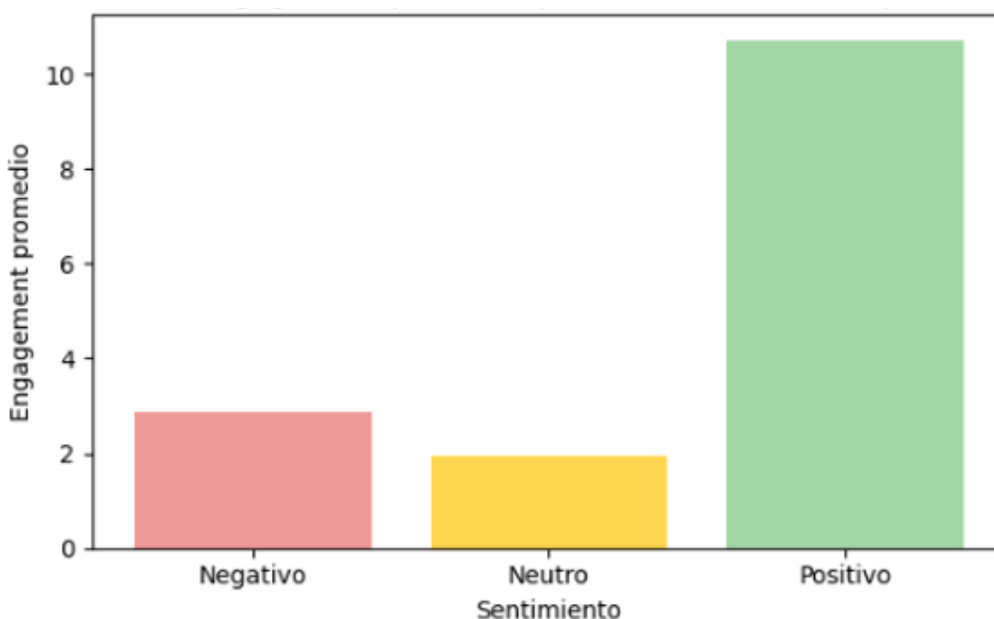


*Nota.* El gráfico presenta el promedio de interacción generado por las publicaciones de la DIAN según su clasificación en sentimiento negativo, neutro y positivo.

Según la Figura 8, en el caso del banco de la república las publicaciones con sentimiento positivo presentan el engagement promedio más alto ( $\approx 10,71$ ), mientras que las negativas ( $\approx 2,87$ ) y las neutras ( $\approx 1,96$ ) registran niveles considerablemente menores. Esta diferencia muestra que ciertos contenidos, independientemente del tono emocional, generan una interacción significativamente mayor por parte del público. Identificar estas publicaciones con alto engagement resulta fundamental para que las entidades decidan de manera estratégica cuáles responder, priorizando aquellas que tienen mayor potencial para influir en la conversación digital y mejorar la percepción ciudadana mediante intervenciones visibles y oportunas.

### Figura 8

*Engagement Promedio por Sentimiento en el Banco de la República*



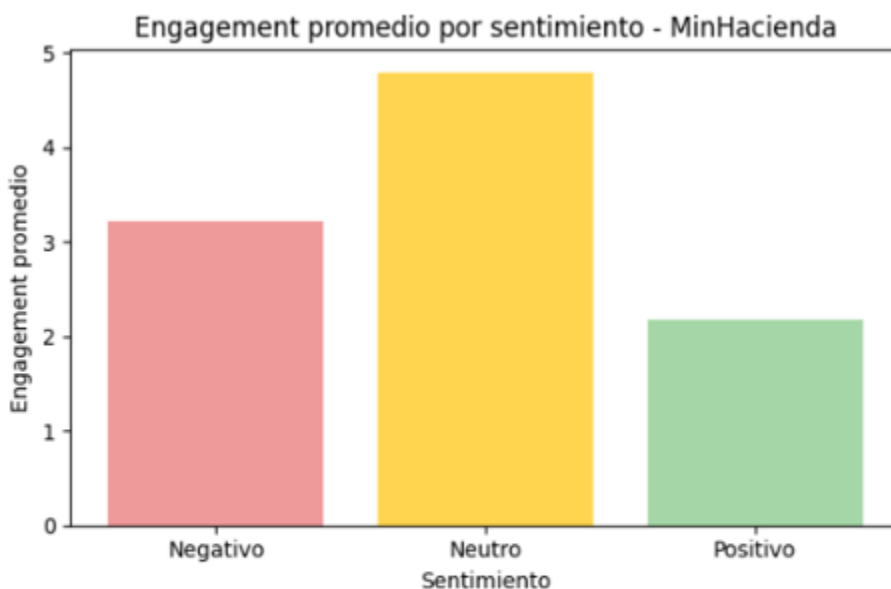
Sentimiento	mean	count
0 Negativo	2.87	150
1 Neutro	1.96	92
2 Positivo	10.71	7

*Nota.* El gráfico presenta el promedio de interacción generado por las publicaciones del Banco de la República según su clasificación en sentimiento negativo, neutro y positivo.

En el caso del Ministerio de Hacienda, como muestra la figura 9, las publicaciones clasificadas con sentimiento neutro concentran el nivel más alto de interacción promedio, superando tanto a las publicaciones negativas como a las positivas. Esta distribución evidencia que la respuesta del público varía según el tipo de sentimiento asociado a cada publicación, lo que permite identificar cuáles grupos generan una mayor participación.

### Figura 9

*Engagement Promedio por Sentimiento en el Ministerio de Hacienda*



```

=== Engagement promedio y número de publicaciones - MinHacienda ===
Sentimiento mean count Entidad
0 Negativo 3.22 95 MinHacienda
1 Neutro 4.79 76 MinHacienda
2 Positivo 2.17 12 MinHacienda

```

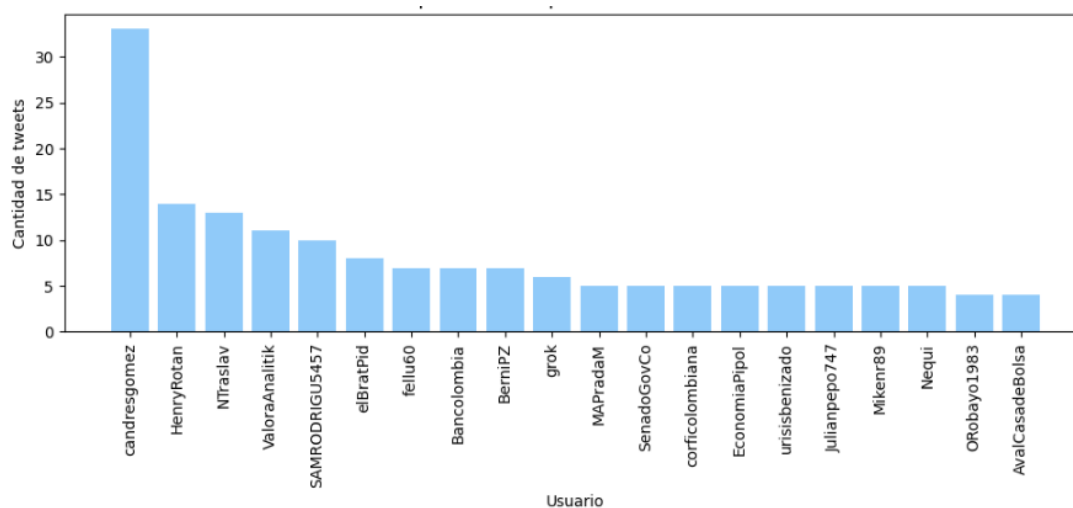
Resulta especialmente interesante observar que la variación del *engagement* entre los distintos tipos de sentimiento es menos marcada que en otras entidades analizadas. Esta menor dispersión sugiere que, independientemente del tono emocional de las publicaciones, la interacción del público tiende a mantenerse en rangos más homogéneos. Esta característica

ofrece un punto de partida valioso para la gestión comunicativa institucional, ya que permite comprender cómo responde la audiencia ante contenidos con carga emocional distinta.

La Figura 10 presenta el comportamiento de los principales emisores dentro del conjunto de datos, permitiendo observar cómo se distribuye el volumen de publicaciones entre los usuarios más activos. Esta visualización resulta clave para comprender qué cuentas impulsan de manera más decidida la conversación en la red social analizada.

### Figura 10

#### *Top 20 Usuarios por Número de Tweets*



*Nota.* El gráfico muestra los veinte usuarios con mayor volumen de publicaciones en el conjunto de datos

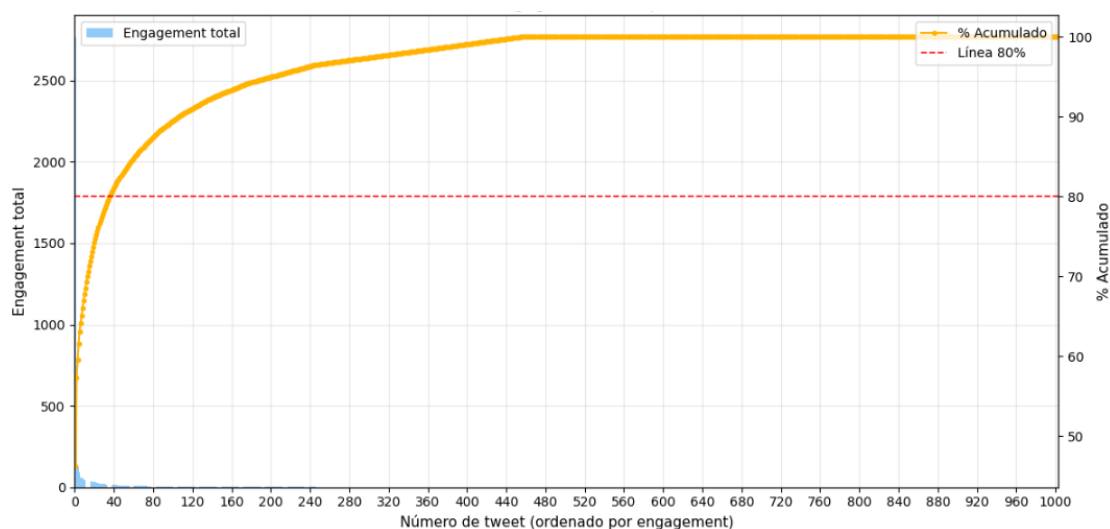
El análisis muestra que la actividad está fuertemente concentrada en un grupo reducido de emisores que publican de forma considerablemente más frecuente que el resto. Esta concentración sugiere que la dinámica de difusión y la visibilidad de ciertos temas pueden estar influenciadas de manera desproporcionada por estas pocas cuentas altamente activas. Identificar

quiénes son estos actores permite contextualizar mejor el flujo informativo y valorar el impacto real que tienen en la configuración de la percepción pública.

No obstante, lo más revelador es la curva de Pareto: solo 35 tweets ( $\approx 3,5\%$  del total) concentran  $\approx 80\%$  del engagement del período. El promedio global de engagement por tweet es  $\approx 5,96$ , pero la atención se distribuye de forma altamente desigual.

### Figura 11

#### *Pareto del Engagement Total por Tweet*



*Nota.* El diagrama muestra que cantidad de tweets acumulan el 80% del engagement total en el conjunto de datos.

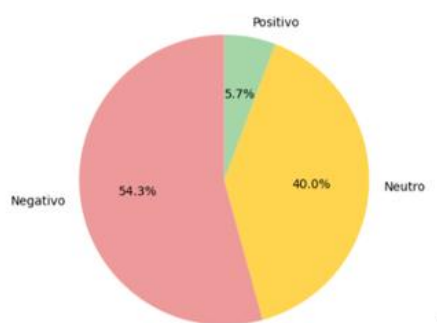
Este hallazgo respalda dos acciones: (i) identificar y aprender de los formatos, timings y recursos de los 35 de mayor tracción; (ii) optimizar pauta y distribución para que los contenidos de servicio clave (especialmente neutros de alto valor informativo) escalen su alcance sin depender de la aleatoriedad algorítmica.

La Figura 12 evidencia que, dentro de los 35 mensajes con mayor interacción, los contenidos negativos constituyen la mayoría, seguidos por los neutros y, en una proporción

mínima, los positivos, lo que revela un patrón donde los picos de engagement tienden a surgir en torno a publicaciones críticas o controversiales. Aunque en promedio el sentimiento neutro presenta niveles más altos de interacción, es el contenido negativo el que se posiciona con mayor fuerza entre los mensajes más visibles, sugiriendo que la atención del público se concentra especialmente en situaciones de tensión, debate o cuestionamiento. Esta dinámica puede estar relacionada tanto con el comportamiento propio de las audiencias digitales —que suelen amplificar más aquello que genera discusión o inconformidad— como con la influencia de cuentas con amplio alcance o con coyunturas de alto interés público que favorecen la viralización selectiva de ciertos temas.

## Figura 12

*Proporción de Sentimientos en el Top 35 de Tweets por Engagement*



	Fecha	Usuario	Sentimiento	engagement_total
87	2025-10-09 12:12:20	Leti_Ariza	Neutro	2763
195	2025-10-10 21:52:04	HELICOOPTERO	Negativo	663
659	2025-10-09 12:00:06	robertoariasr	Negativo	129
930	2025-10-29 22:05:05	RTVCnoticias	Negativo	123
61	2025-10-09 17:44:53	Juli_Zapata_E	Negativo	92
330	2025-10-11 19:05:01	EnLineaBVC	Positivo	61
380	2025-10-11 21:11:01	ORobayo1983	Neutro	57
324	2025-10-11 21:11:01	ORobayo1983	Neutro	57
70	2025-10-09 16:54:18	tomosed33223801	Neutro	55
749	2025-10-06 22:34:06	JD_Quinteros	Neutro	47
646	2025-10-09 13:40:39	senadormotoa	Neutro	46
31	2025-10-09 19:14:30	JUANJIM48591357	Negativo	45
995	2025-10-29 17:53:18	RTVCnoticias	Negativo	42
6	2025-10-09 22:10:59	maferojas	Neutro	40
730	2025-10-08 12:00:11	SenalColombia	Neutro	39
783	2025-10-06 18:31:13	lavozdepueblo1	Negativo	36
459	2025-10-12 20:27:01	alfredosaadev	Neutro	35
279	2025-10-07 00:37:03	PoliciaAduanera	Neutro	35
764	2025-10-06 20:39:59	heidy_up	Neutro	34
653	2025-10-09 12:21:58	JuanSlopezH	Negativo	34
113	2025-10-11 23:44:20	Vendetta1313_	Negativo	31
15	2025-10-09 20:56:20	Alejandro8220	Neutro	29
409	2025-10-13 20:05:33	e18ratPid	Neutro	29
164	2025-10-11 15:43:23	feilu60	Negativo	27
11	2025-10-09 21:39:38	feilu60	Negativo	24
631	2025-10-09 15:22:33	FEDe_Colombia_	Negativo	24
975	2025-10-29 18:43:28	deusmercurius	Negativo	23
386	2025-10-14 01:20:19	exillum08	Negativo	20
176	2025-10-11 04:29:25	MikyRodriguez01	Negativo	20
27	2025-10-09 19:38:00	Guando2005	Negativo	19
957	2025-10-29 19:38:07	MelinaCeronR	Negativo	19
184	2025-10-11 00:25:46	angiesalas9	Positivo	18
718	2025-10-08 13:43:23	MiNHacienda	Neutro	18
254	2025-10-07 05:51:20	CabalPaloma	Negativo	18
305	2025-10-12 14:02:51	ValoraAnalitik	Negativo	18

Esto refuerza la necesidad de monitoreo en tiempo real y respuesta temprana ante temas que escalan rápidamente, además de empaquetar información de alto valor en formatos que faciliten compartibilidad (infografías, hilos explicativos, video corto)

Se realiza un análisis de los términos más frecuentes utilizados en las publicaciones, iniciando por aquellos clasificados con sentimiento positivo. La Figura 13 permite visualizar con claridad cuáles son las palabras que predominan en este conjunto, revelando un campo semántico asociado a temas institucionales, desarrollo económico y construcción colectiva. En esta categoría sobresalen conceptos como “economía”, “trabajo”, “país”, “fortalecer” y “compromiso”, que reflejan una narrativa orientada a destacar acciones, esfuerzos y objetivos comunes dentro del discurso positivo.

### Figura 13

*Nube de Palabras para el Sentimiento Positivo*



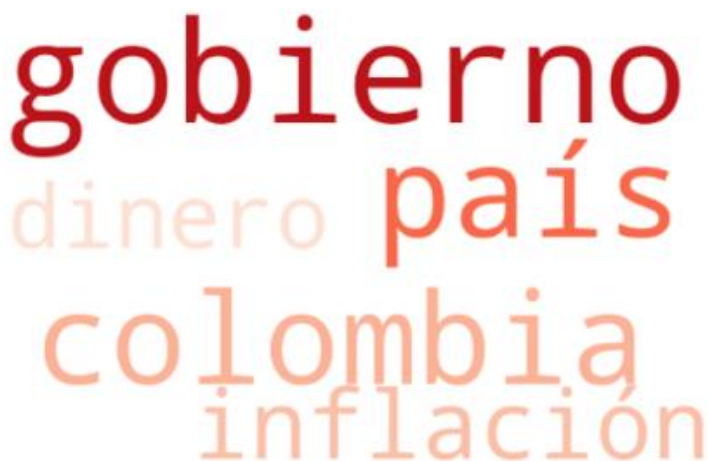
```
=== Términos frecuentes POSITIVO ===
    termino  frecuencia
0   economía      8
1    trabajo      5
2     país        5
3  fortalecer     3
4  compromiso     3
```

*Nota.* Se muestran las 5 palabras que más se repiten en tweets con sentimiento positivo

En cuanto al sentimiento negativo, la figura 14 muestra que el análisis de términos más frecuentes revela un conjunto de palabras estrechamente ligado a asuntos políticos y económicos, lo que indica que las percepciones desfavorables se articulan en torno a temas de alta sensibilidad pública. En esta categoría sobresalen términos como “gobierno”, “Colombia”, “país”, “inflación” y “dinero”, que estructuran el núcleo del discurso crítico dentro del conjunto de datos. Resulta particularmente interesante que la palabra “país” aparezca tanto en los mensajes positivos como en los negativos, lo que sugiere que, independientemente de la valoración emocional, existe una preocupación constante por temas nacionales que atraviesan la conversación digital. Esta coincidencia refuerza la idea de que el debate público —sea positivo o negativo— se enmarca en una narrativa colectiva centrada en el rumbo y la situación del país.

#### Figura 14

*Nube de Palabras para el Sentimiento Negativo*



```

=== Términos frecuentes NEGATIVO ===
      termino  frecuencia
0  gobierno      39
1  colombia      31
2    país        31
3  inflación     31
4    dinero      27

```

*Nota.* Se muestran las 5 palabras que más se repiten en tweets con sentimiento negativo

La Figura 15 muestra que, en el caso del sentimiento neutro, los términos más frecuentes se agrupan alrededor de referencias institucionales y temas de alcance nacional. Entre las palabras que predominan se encuentran “Colombia”, “Dian”, “banco”, “presidente” e “inflación”, lo que evidencia un enfoque discursivo orientado hacia entidades públicas, asuntos económicos y actores gubernamentales

### Figura 15

*Nube de Palabras para el Sentimiento Negativo*



```

=== Términos frecuentes NEUTRO ===
      termino  frecuencia
0     colombia         24
1         dian         18
2         banco         15
3     presidente         14
4     inflación         14

```

*Nota.* Se muestran las 5 palabras que más se repiten en tweets con sentimiento neutro

## Conclusiones

Se logró consolidar un corpus de 1.000 publicaciones provenientes de la red social X, recolectadas entre el 6 y el 29 de octubre de 2025. La extracción, efectuada mediante la API de X y la librería requests en Python, permitió obtener mensajes asociados a las entidades DIAN, Ministerio de Hacienda y Banco de la República. El conjunto de datos resultante cumplió con el propósito de capturar la conversación pública sobre temas tributarios, aduaneros y cambiarios, garantizando una muestra suficiente y representativa para el análisis posterior.

El proceso de limpieza y normalización de texto, implementado con pandas y expresiones regulares, permitió eliminar ruido informativo —como URLs, menciones, hashtags y duplicados—, estandarizar codificaciones y asegurar la coherencia de los datos. Esta depuración mejoró la calidad del corpus y redujo el sesgo introducido por caracteres no textuales o ruido social propio de la plataforma, facilitando la aplicación fiable de técnicas de análisis de sentimiento en español.

La clasificación de sentimientos mediante pysentimiento mostró una predominancia del tono negativo (58%), seguido del neutro (37%) y un bajo porcentaje de positivo (4%). Este patrón evidencia que la percepción ciudadana hacia el sistema tributario, aduanero y cambiario tiende a ser crítica, particularmente en torno a la DIAN, que concentró la mayoría de las menciones negativas. Se comprobó así la utilidad del análisis de sentimiento automatizado como herramienta para identificar patrones emocionales en el discurso digital sobre entidades públicas.

El examen de frecuencias y nubes de palabras reveló diferencias claras en el lenguaje asociado a cada tipo de sentimiento: términos como “gobierno”, “inflación” y “dinero” destacaron en los mensajes negativos, mientras que “economía”, “trabajo” y “fortalecer” fueron más frecuentes en los positivos. A su vez, el análisis de engagement mostró que los

contenidos neutros generan mayor interacción promedio, y que un reducido grupo de usuarios concentra la conversación. Estos hallazgos aportan evidencia empírica para comprender la dinámica comunicativa en redes sociales y su relación con la percepción institucional.

El análisis Pareto mostró que solo 35 publicaciones concentran cerca del 80% del engagement total, lo que indica que la conversación digital se estructura alrededor de un conjunto reducido de mensajes altamente influyentes. Este resultado sugiere que, al identificar esas publicaciones de mayor impacto, las entidades públicas pueden dirigir su atención hacia los puntos donde realmente se forma la percepción ciudadana. Responder, aclarar o complementar la información contenida en esos tweets de alto alcance permitiría que las entidades intervengan de manera estratégica en los espacios de mayor visibilidad, aumentando la probabilidad de generar un efecto real y positivo en la percepción pública sobre sus acciones y decisiones

### Referencias Bibliográficas

- AbuRaed, A. G. T. (2024). Topic modeling and sentiment analysis of Twitter data: A case study on public perception of COVID-19 measures. *JMIR Formative Research*, 8(1), e59425. <https://doi.org/10.2196/59425>
- Antypas, D., Preece, A., & Camacho-Collados, J. (2022). *Negativity spreads faster: A large-scale multilingual sentiment analysis on political tweets*. *arXiv preprint arXiv:2202.00396*. <https://doi.org/10.48550/arXiv.2202.00396>
- Arango Espinal, E., Osorio Andrade, C. F., & Uribe Jiménez, M. (2025). Eficacia del gobierno y su relación con el respaldo ciudadano en redes sociales: El caso del Paisaje Cultural Cafetero de Colombia. *Estudios Políticos*, (72). <https://doi.org/10.17533/udea.espo.n72a11>
- Camacho-Collados, J., Rezaee, K., & Riahi, T. (2022). TweetNLP: Cutting-edge natural language processing for social media. *arXiv preprint arXiv:2206.14774*. <https://doi.org/10.48550/arXiv.2206.14774>
- Conti, D., Gómez, C. E., Jaramillo, J. G., & Ospina, V. E. (2023). Monitoring the quality and perception of service in Colombian public service companies with Twitter and descriptive temporal analysis. *Applied Sciences*, 13(18), 10338. <https://doi.org/10.3390/app131810338>
- Duran-Vaca, M. K., & Ballesteros-Ricaurte, J. A. (2020). Sentiment analysis on Twitter to measure the perception of taxation in Colombia. *ResearchGate*. [https://doi.org/10.1007/978-3-030-36778-7\\_20](https://doi.org/10.1007/978-3-030-36778-7_20)
- Figueroa Perdomo, F. A. (2020). Las redes sociales como mecanismo de participación ciudadana en la gestión pública y sus perspectivas futuras en Colombia [Trabajo de grado,

Universidad Nacional Abierta y a Distancia]. *Repositorio Institucional UNAD*.

<https://repository.unad.edu.co/handle/10596/36517>

Iguarán Fernández, J. J., Pérez, J. M., & Rosati, G. (2024). Identification of emotions on Twitter during the 2022 electoral process in Colombia. arXiv preprint arXiv:2407.07258.

<https://doi.org/10.48550/arXiv.2407.07258>

Joloudari, J. H., Hosseinian-Far, A., Taghiyareh, F., & Saad, E. M. (2022). BERT–Deep CNN: State-of-the-art sentiment analysis model for sentiment analysis of COVID-19 tweets.

arXiv preprint arXiv:2211.09733. <https://doi.org/10.48550/arXiv.2211.09733>

Kaschesky, M., Sobkowicz, P., & Bouchard, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4), 470–479. <https://doi.org/10.1145/2037556.2037607>

Mat Razali, N. A., Malizan, N. A., Hasbullah, N. A., Wook, M., Mohd Zainuddin, N., Ishak, K. K., Ramli, S., & Sukardi, S. (2021). Opinion mining for national security: Techniques, domain applications, challenges and research opportunities. *Journal of Big Data*, 8(1), 150. <https://doi.org/10.1186/s40537-021-00536-5>

Moreno-Sandoval, L. G., López-Herrera, A. G., & Rodríguez-González, M. (2024). Detection of hate speech, racism, and misogyny in Spanish social media using transformer-based models. *Big Data and Cognitive Computing*, 8(9), 113.

<https://doi.org/10.3390/bdcc8090113>

Quinchía-Saavedra, C., Castañeda García, J. A., & Rey-Pino, J. M. (2025). Tweeting for peace: An analysis of Twitter use in Colombia's peace process. *International Review on Public and Nonprofit Marketing*. <https://doi.org/10.1007/s12208-025-00433-0>

- Ruiz Soto, M. A. (2023). Estado del arte de la comunicación de gobierno en tiempos de medios sociales. *Ópera*, (32), 81–107. <https://doi.org/10.18601/16578651.n32.06>
- Russell, S. N., Rao-Graham, L., & McNaughton, M. (2024). Mining social media data to inform public health policies: A sentiment analysis case study. *Revista Panamericana de Salud Pública*, 48, e79. <https://doi.org/10.26633/RPSP.2024.79>
- Suhaimin, M. S., Hijazi, M. H. A., Mounq, E. G., Nohuddin, P. N. E., Chua, S., & Coenen, F. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2023.101776>
- Venkit, P. N., & Wilson, S. (2021). Identifying bias against people with disabilities in sentiment analysis and toxicity detection models. arXiv preprint arXiv:2111.13259. <https://doi.org/10.48550/arXiv.2111.13259>
- Vizmanos, J. F. V., Siar, S. V., Albert, J. R. G., Sarmiento, J. L. C., & Hernandez, A. C. (2023). Like, comment, and share: Analyzing public sentiments of government policies in social media (PIDS Discussion Paper Series No. 2023-33). *Philippine Institute for Development Studies*. <https://doi.org/10.62986/dp2023.33>