



## **COVID-19 EN REDES SOCIALES: ANÁLISIS DE SENTIMIENTO EN SUDAMÉRICA, 2020 \***

**Pilar Villena Guzmán**  
pilarvillena@edu.uah.es

**Jillie Chang Kcomt**  
jillie.chang@edu.uah.es

**Universidad Alcalá de Henares (UAH) Madrid**  
**Máster en Business Intelligence y Data Science**  
Asesor: Dr. Lino González García

# Resumen

1. Introducción
2. Datos
3. Modelos
4. Resultados
5. Conclusiones

## Contextualización

- La enfermedad por coronavirus (COVID-19) reportada en diciembre de 2019 ha generado no solo una crisis sanitaria y económica, sino también una crisis social de miedo masivo y fenómenos de pánico que han afectado a la población.
- Esto genera que sea importante medir el sentimiento de la población de modo que los Gobiernos puedan transmitir mensajes apropiados y oportunos a sus ciudadanos. Asimismo, es importante que el Gobierno escuche las opiniones sobre las políticas implementadas a modo de extraer retroalimentación que busque la mejora continua.

## Definición de Análisis de sentimientos

- Una técnica utilizada para observar las sensaciones generadas en la población respecto al COVID-19 es el Análisis de sentimientos.

*El Análisis de Sentimiento (minería de opinión): uso del Procesamiento de Lenguaje Natural (PLN) para determinar automáticamente el sentimiento que una persona está expresando en un extracto de texto. Este sentimiento puede ser clasificado de manera binaria, terciaria, o con múltiples categorías.*

Zhang, M., Ng, J. (2020)

- Existen diferentes métodos para clasificar los sentimientos que pueden ser agrupados en aprendizaje supervisado y no supervisado

## Revisión de literatura

**Cuadro 1.** Literatura revisada

Autor	Lugar	Principales resultados	Método
Samuel, J. et ál. (2020)	Estados Unidos	Sentimientos negativos y miedo se incrementan a medida que los contagio aumentaba	A. supervisado
Barkur, G. et ál. (2020)	India	Impresión positiva sobre medidas tomadas por el gobierno (sentimientos de molestia y preocupación ante la demora del establecimiento de cuarentena)	A. no supervisado
Dubey, A. D. (2020)	12 países europeos	Sentimientos positivos y esperanzadores, aunque con instantes de tristeza y preocupación	
Sharma K et ál. (2020)	20 países (inglés)	Impresión positiva sobre medidas de relacionadas con el teletrabajo y distanciamiento social	
Zhang, M., Ng, J. (2020)	Reino Unido	impresión más positiva luego cuarentena. Los tweets que contenían "gobierno" eran más negativos antes de la cuarentena. Hay una reacción más positiva a "quédate en casa" que a "cuarentena"	A. supervisado y no supervisado

# Objetivos

- En ese contexto, este trabajo de investigación tiene los siguientes objetivos:
  - i. Medir y comparar el sentimiento de los ciudadanos en Sudamérica con respecto a la COVID-19 y a las medidas tomadas por sus gobiernos.
  - ii. Identificar cuáles han sido los temas de interés durante las medidas adoptadas ante la COVID-19.

## Extracción de información

- Fuente: Twitter (402 229 tweets )**

- ✓ *Periodo de descarga: 1 de enero hasta el 31 de agosto de 2020.*
- ✓ *Zona de descarga: capital de Perú, Uruguay, Ecuador, Argentina, Colombia y Chile*
- ✓ *Palabras claves: "covid", "pandemia", "coronavirus" y "cuarentena"*
- ✓ *Idioma: español*

- Método: Librerías de Python**

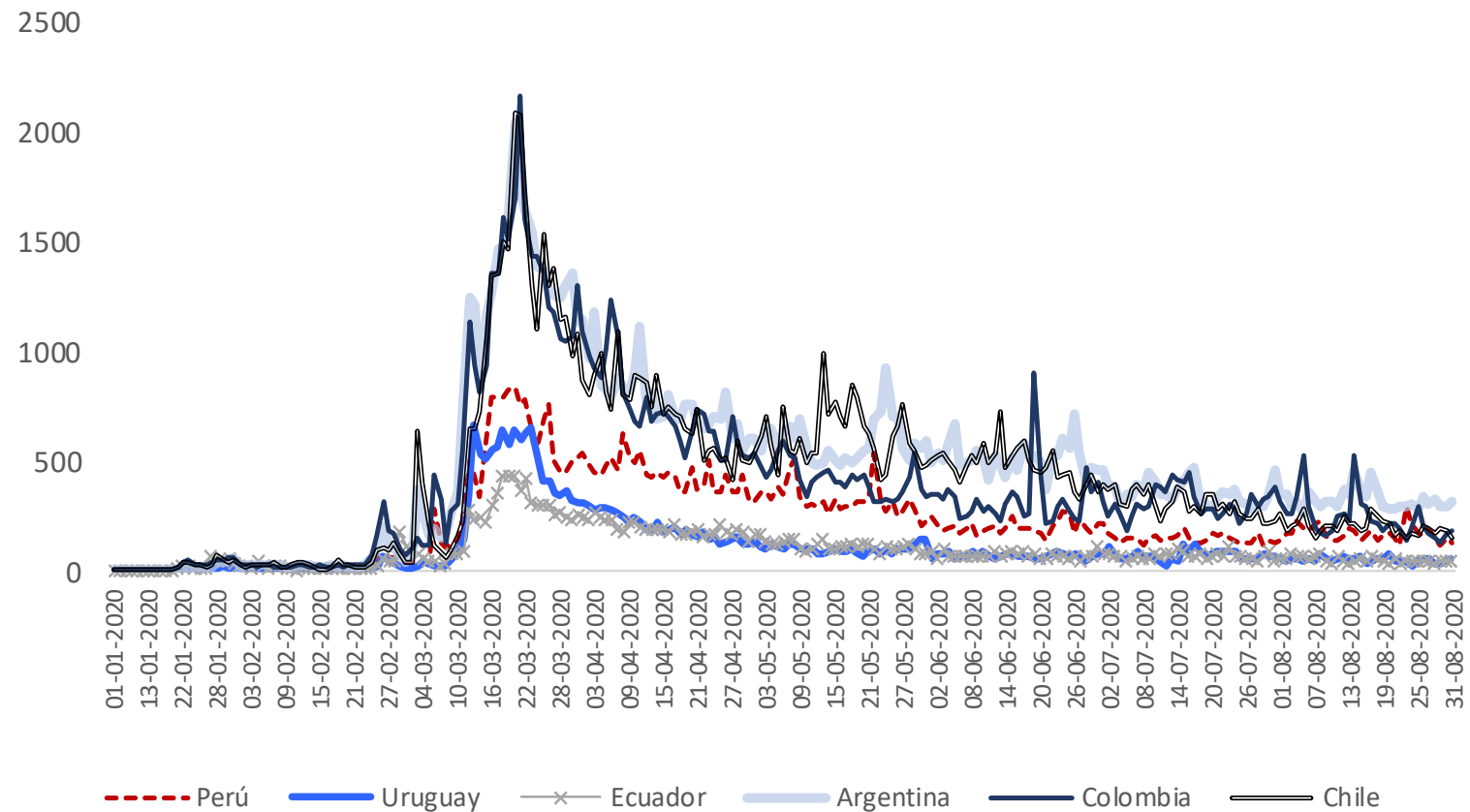
- ✓ *GetOldTweets3*
- ✓ *Tweepy*

**Cuadro 2.** Variables que conforman la base de datos del estudio

Librería	Variable	Descripción
GetOldTweets3	Tweet_Id	Identificador del Tweet
	Tweet_User_Id	Identificador del Usuario
	Text	Texto
	Datetime	Fecha del tweet
	Hashtags	Etiquetas
	Location	Capital del país identificado a partir del "geocode" y "distances"
Tweepy	Tweet_Source	Fuente de origen de tweet
	lang	Idioma

## Contraste de información extraída con hechos reales

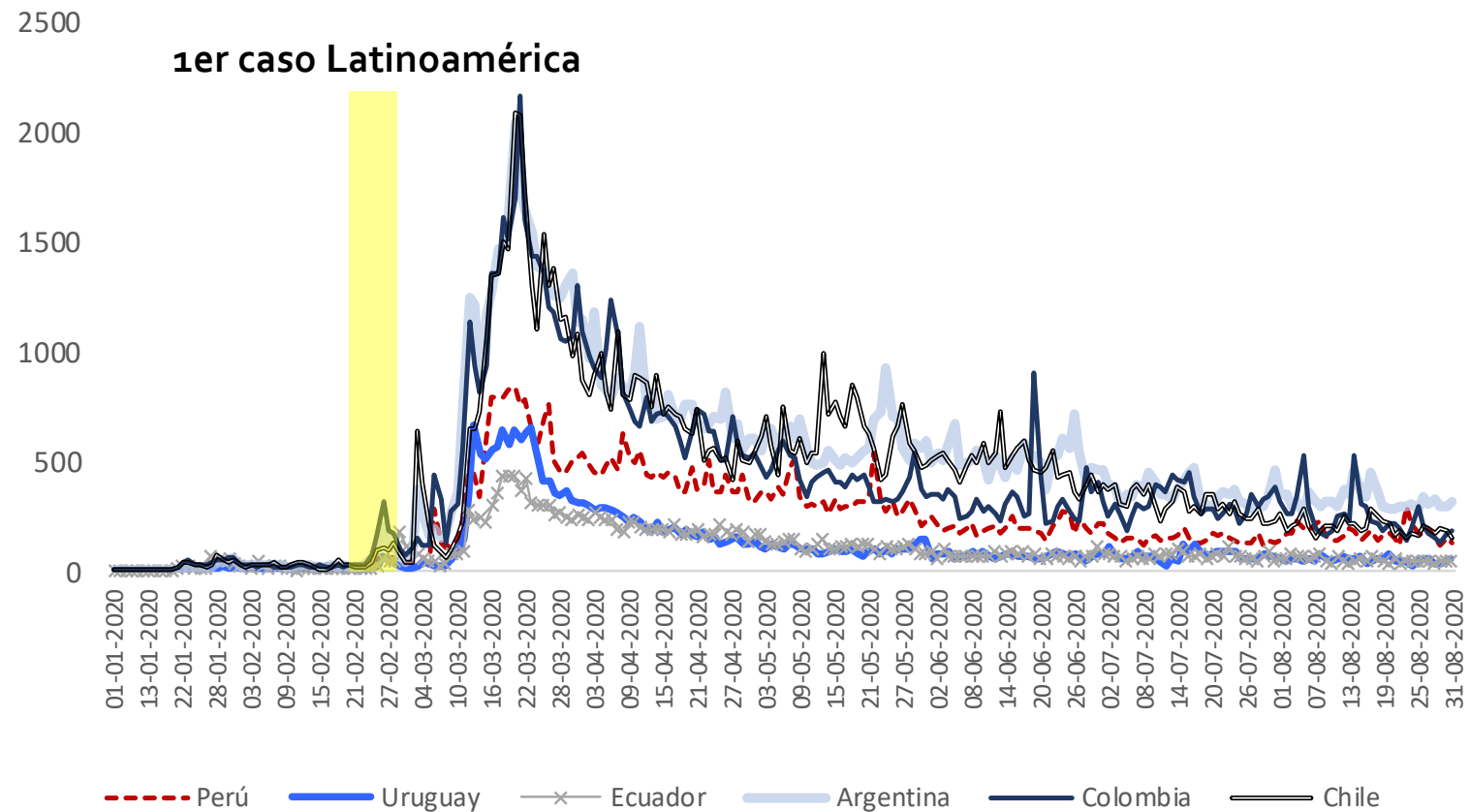
**Gráfico 1.** Evolución de la cantidad de tweets por país, 01-01-2020 al 31-08-2020





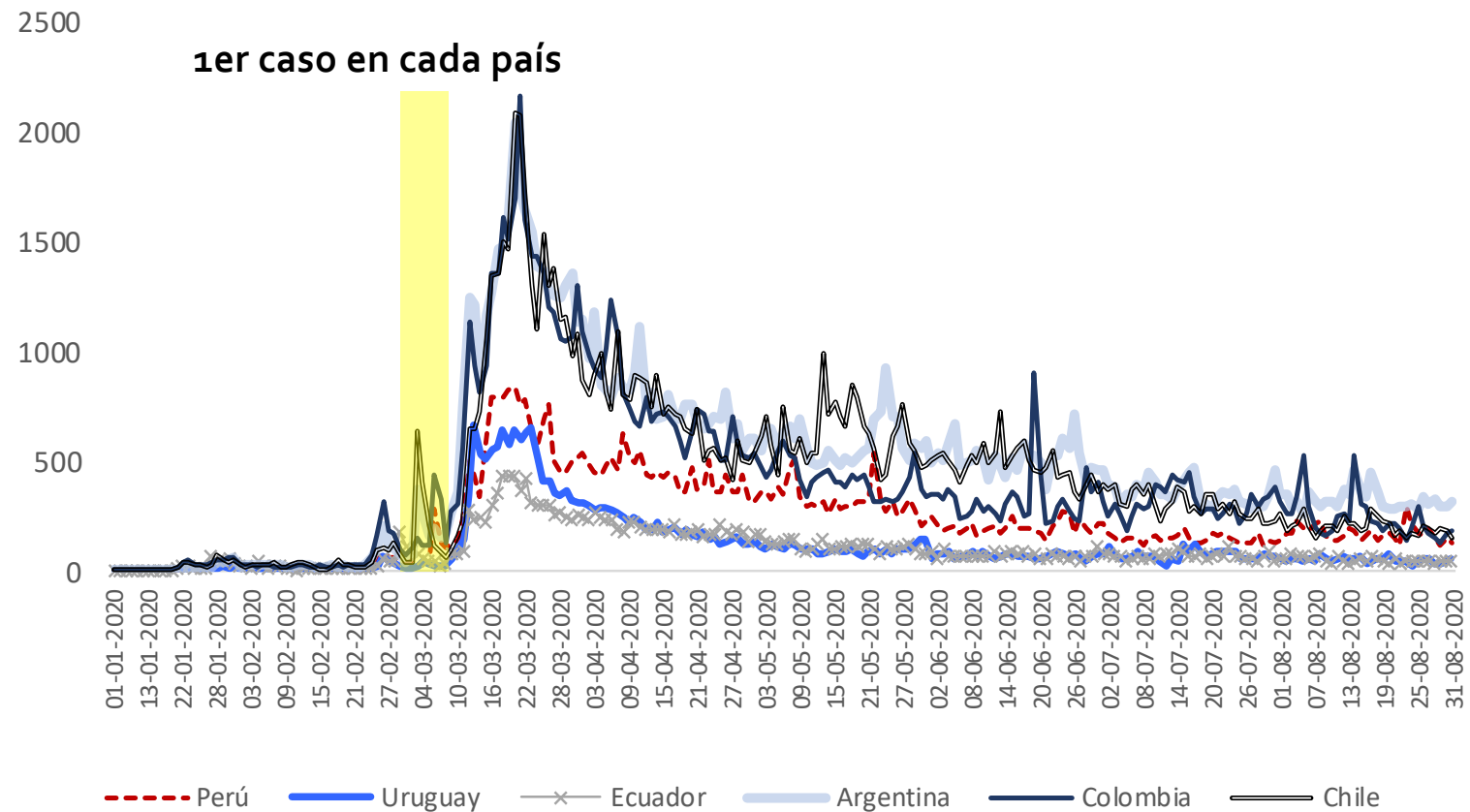
# Contraste de información extraída con hechos reales

**Gráfico 1.** Evolución de la cantidad de tweets por país, 01-01-2020 al 31-08-2020



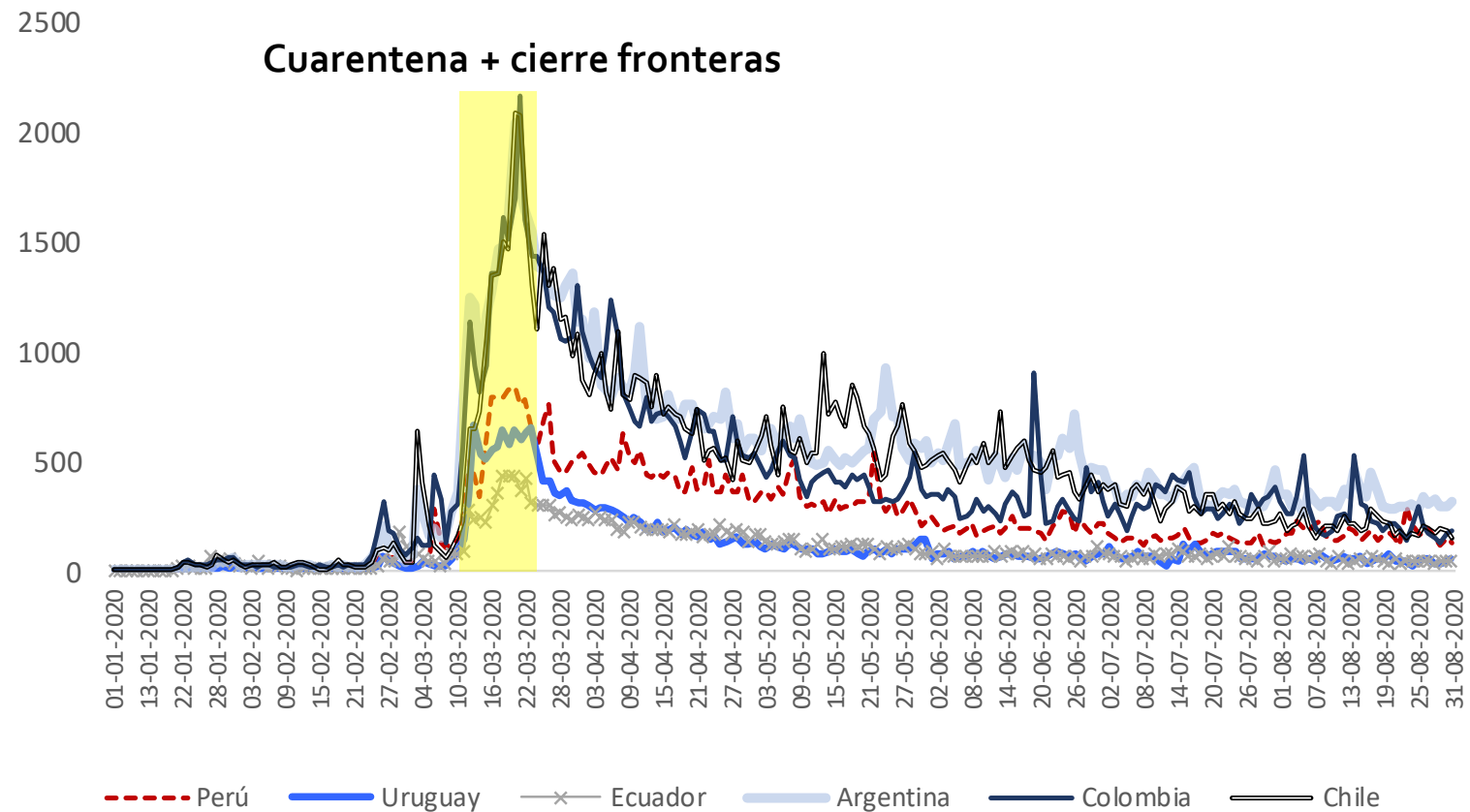
# Contraste de información extraída con hechos reales

**Gráfico 1.** Evolución de la cantidad de tweets por país, 01-01-2020 al 31-08-2020



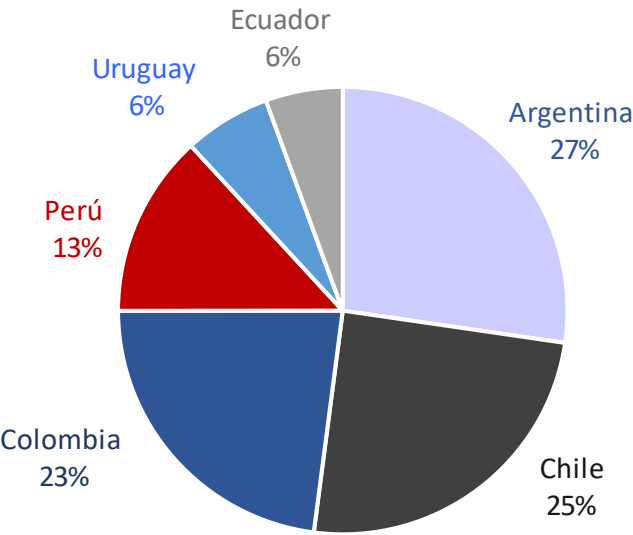
# Contraste de información extraída con hechos reales

**Gráfico 1.** Evolución de la cantidad de tweets por país, 01-01-2020 al 31-08-2020



# Análisis descriptivo de los datos

**Gráfico 2.** Distribución de cantidad de tweets por país

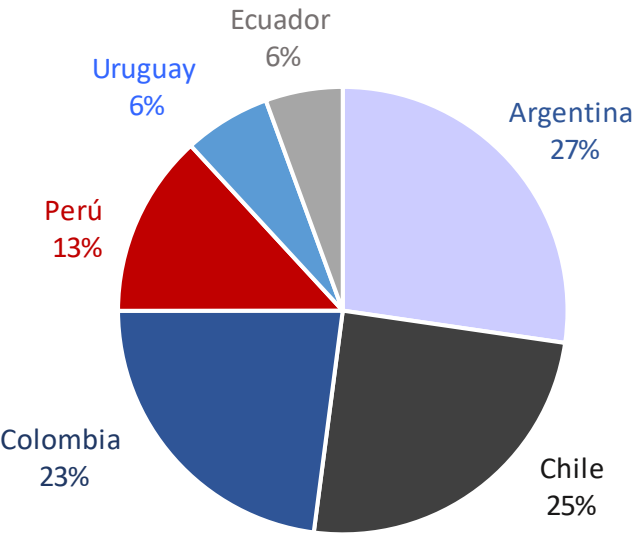


**Cuadro 3.** Principales hashtags por país

País	Principales Hashtags
Argentina	coronavirus, cuarentena, quedateencasa, yomequedoencasa, covid19, argentina, pandemia, coronavirusargentina, buenos aires
Chile	coronavirus, covid_19, cuarentena, quedateencasa, chile, pandemia, cuarentenatotal, covid19Chile, coronavirusenchile
Colombia	coronavirus, covid_19, cuarentena, colombia, quedateencasa, yomequedoencasa, bogota, pandemia, coronavirusencolombia
Perú	cuarentena, coronavirus, yomequedoencasa, covid_19, quedateencasa, peru, lima, pandemia, coronavirusperu
Uruguay	coronavirus, covid_19, cuarentena, quedateencasa, uruguay, coronavirusenuruguay, yomequedoencasa, montevideo
Ecuador	covid_19, coronavirus, ecuador, cuarentena, quedateencasa, quito, covid_19ec, urgente, yomequedoencasa

# Análisis descriptivo de los datos

**Gráfico 2.** Distribución de cantidad de tweets por país

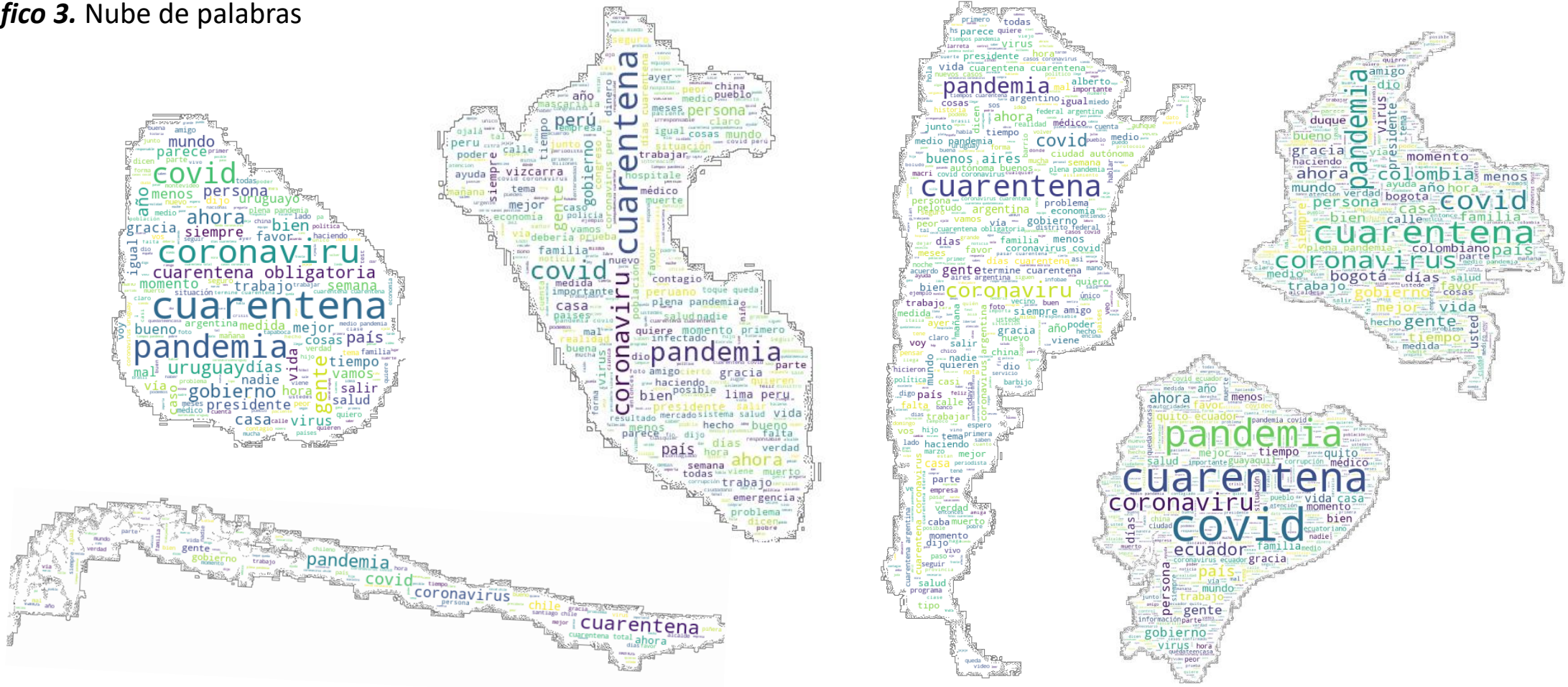


**Cuadro 3.** Distribución de principales hashtags por país

País	Principales Hashtags
Argentina	coronavirus, cuarentena, <b>quedateencasa</b> , <b>yomequedoencasa</b> , covid19, argentina, pandemia,coronavirusargentina,buenos aires
Chile	coronavirus, covid_19, cuarentena, <b>quedateencasa</b> , chile,pandemia, cuarentenatotal, covid19Chile, coronavirusenchile
Colombia	coronavirus, covid_19, cuarentena, colombia, <b>quedateencasa</b> , <b>yomequedoencasa</b> , bogota, pandemia, coronavirusencolombia
Perú	cuarentena, coronavirus, <b>yomequedoencasa</b> , covid_19, <b>quedateencasa</b> , peru, lima, pandemia, coronavirusperu
Uruguay	coronavirus, covid_19, cuarentena, <b>quedateencasa</b> , uruguay, coronavirusenuruguay, <b>yomequedoencasa</b> , montevideo
Ecuador	covid_19, coronavirus, ecuador, cuarentena, <b>quedateencasa</b> , quito, covid_19ec, urgente, <b>yomequedoencasa</b>

# Análisis de textos

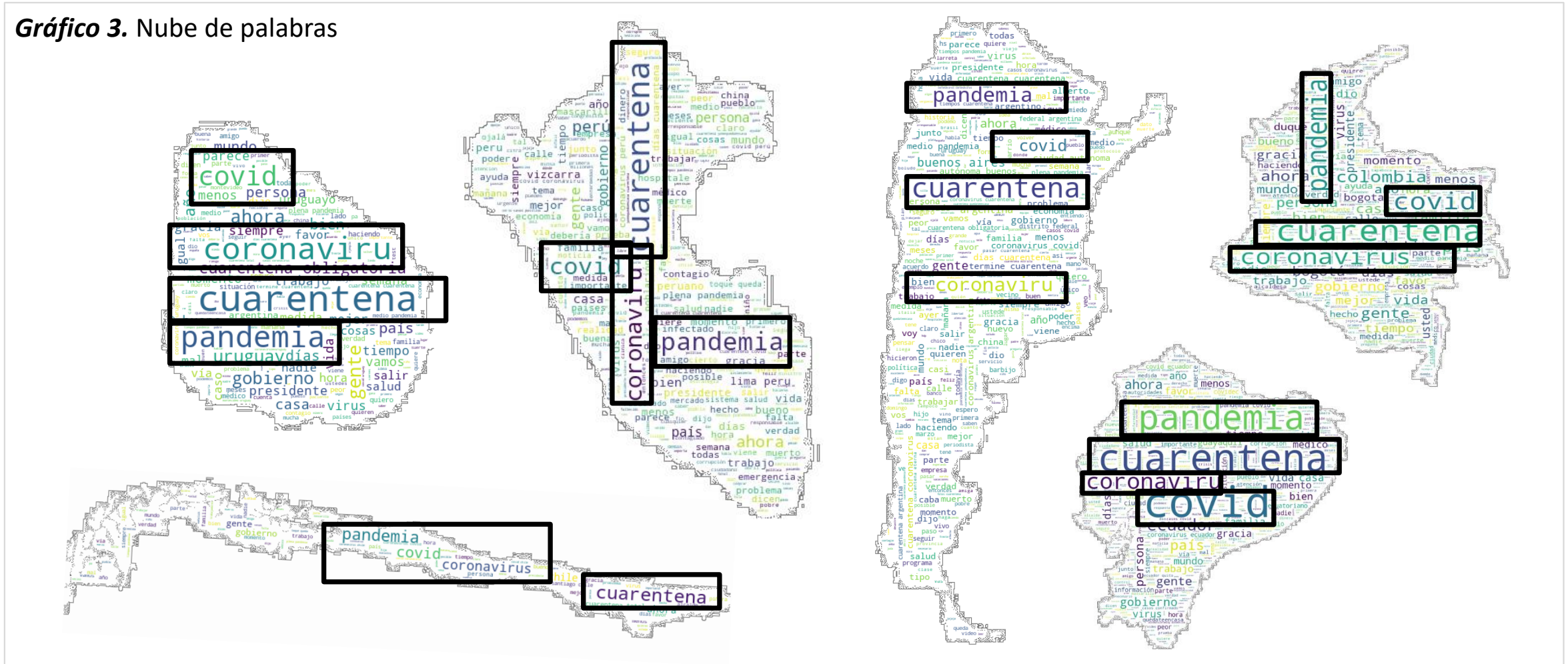
**Gráfico 3.** Nube de palabras





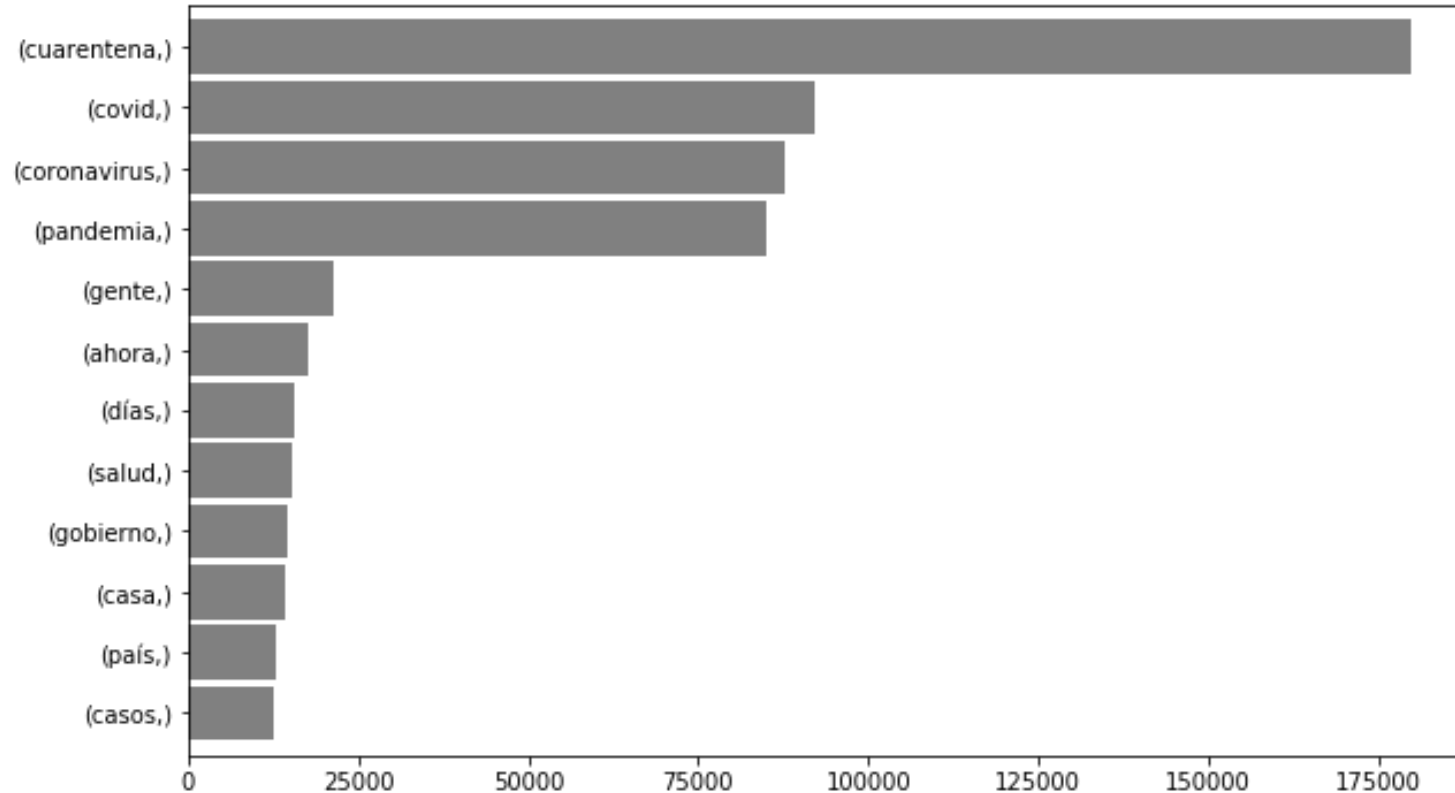
# Análisis de textos

**Gráfico 3.** Nube de palabras



## Análisis de textos

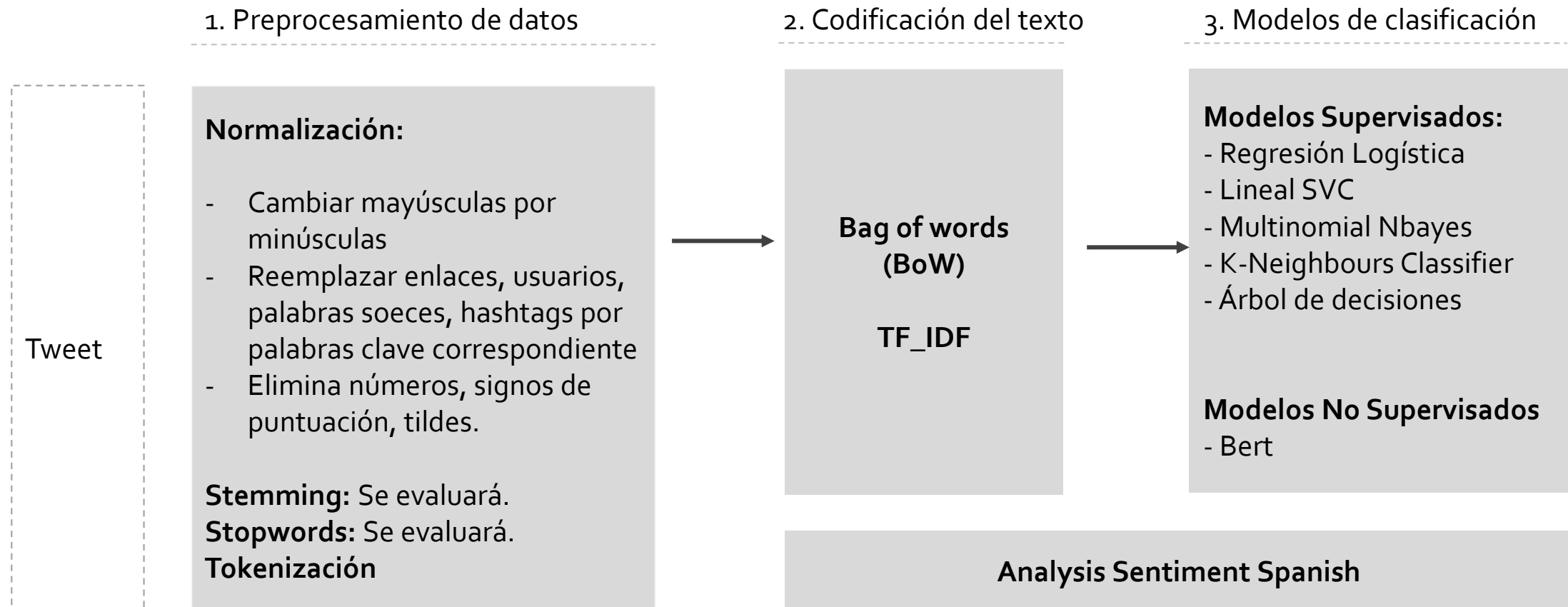
**Gráfico 4.** Unigrama





# Modelos de Análisis de Sentimiento

**Gráfico 5.** Modelos de Análisis de Sentimiento



## Etiquetado mediante emojis

- Se etiquetaron automáticamente como “positivos” aquellos tweets que contenían emojis como 😊 😄 😍 y “negativos” aquellos que contenían 😞, 😟.
- Base de entrenamiento de 22274 tweets, con igual cantidad de tweets positivos y negativos
- Base test de 9546 tweets

## Evaluación del rendimiento del modelo

- AUC: Muestra la capacidad del modelo para distinguir clases de sentimientos positivos y negativos. Valores a partir de 0.7 se consideran aceptables
- F1W: valor que combina las medidas de precisión y exhaustividad en un solo valor.

## Modelos de clasificación

**Cuadro 4a.** Resultados de modelos

Normalización	Stemming	Elimina Stopwords	Extracción de características	Regresión Logística		Lineal SVC		Multinomial Nbayes		Kneighbors Classifier		Árboles de decisión	
				AUC	F1W	AUC	F1W	AUC	F1W	AUC	F1W	AUC	F1W
Sí	No	Sí	BOW	0.76	0.70	0.72	0.67	0.78	0.71	0.65	0.57	0.64	0.63
Sí	No	No	BOW	0.77	0.70	0.73	0.67	0.78	0.71	0.66	0.50	0.63	0.63
Sí	Sí	Sí	BOW	0.77	0.70	0.73	0.67	0.78	0.71	0.66	0.56	0.63	0.63
Sí	Sí	No	BOW	0.77	0.70	0.73	0.68	0.78	0.71	0.67	0.51	0.62	0.62
Sí	No	Sí	IT -FD	0.78	0.71	0.76	0.69	0.78	0.71	0.72	0.67	0.62	0.62
Sí	No	No	IT -FD	0.78	0.71	0.76	0.69	0.79	0.71	0.73	0.68	0.61	0.61
Sí	Sí	Sí	IT -FD	0.78	0.71	0.76	0.69	0.78	0.71	0.73	0.68	0.62	0.62
<b>Sí</b>	<b>Sí</b>	<b>No</b>	<b>IT -FD</b>	<b>0.79</b>	<b>0.72</b>	<b>0.77</b>	<b>0.70</b>	<b>0.79</b>	<b>0.71</b>	<b>0.73</b>	<b>0.68</b>	<b>0.61</b>	<b>0.61</b>

**Cuadro 4b.** Resultados de modelos

Normalización	Stemming	Elimina Stopwords	Sentiment Spanish		Bert	
			AUC	F1W	AUC	F1W
Sí	No	No	0.61	0.46	0.69	0.67
Sí	Sí	No	0.58	0.19	0.67	0.65

## Capacidad predictiva de modelo seleccionado

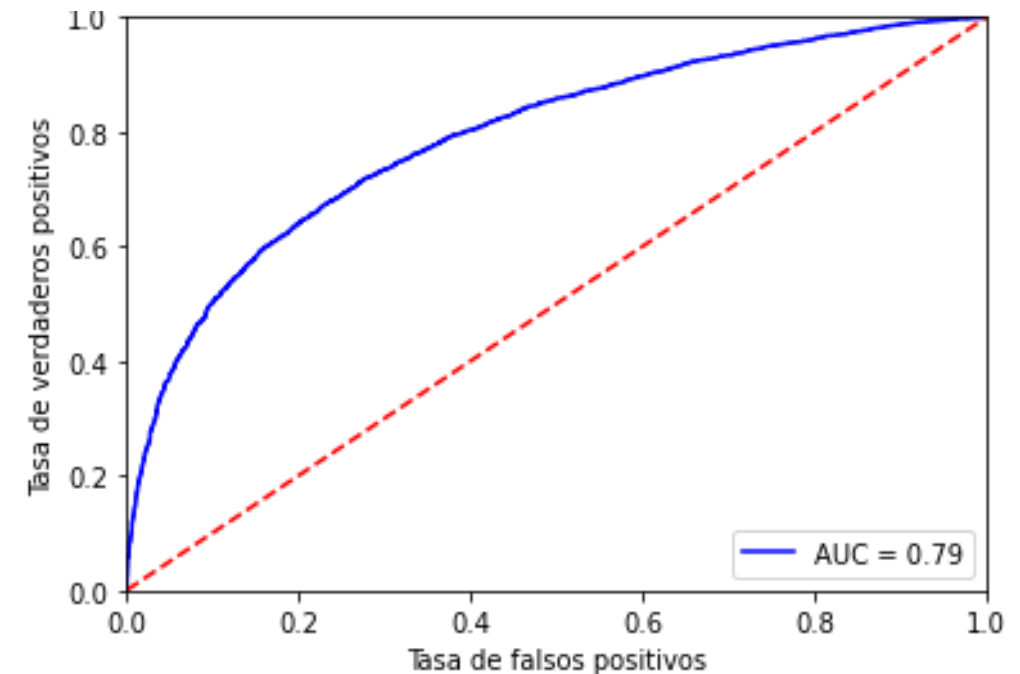
**Cuadro 5.** Matriz de confusión

Matriz de confusión		Polaridad estimada		
		Tweets positivos	Tweets negativos	Total
Polaridad observada	Tweets positivos	3260	1531	4791
	Tweets negativos	1144	3611	4755
	Total	4404	5142	9546

**Cuadro 6.** Indicadores de capacidad predictiva

Indicador	Porcentaje
Sensibilidad/Recall (tasa positiva real)	68.0%
Especificidad (tasa negativa real)	75.9%
Precisión	74.0%
Exactitud (Accuracy)	72.0%
F1-ponderado (F1W)	70.9%

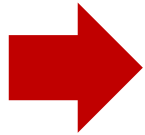
**Gráfico 6.** Curva ROC



- A partir del modelo seleccionado, se predijo el sentimiento con relación a la COVID-19 derivado de cada tweet (se asignó un puntaje de 0-1 a cada texto).
- Se determinó que un tweet es positivo si el puntaje es mayor o igual a 0.6 y negativo en caso contrario.

**Negativo**

puntaje de 0.02  
según el modelo



"Ningún país se librará del virus. Así que dejen de andar diciendo que es histeria colectiva o que es una estrategia por ser potencia mundial #coronavirus"

Tweet del 26 febrero, Colombia

**Positivo**

puntaje de 0.773  
según el modelo

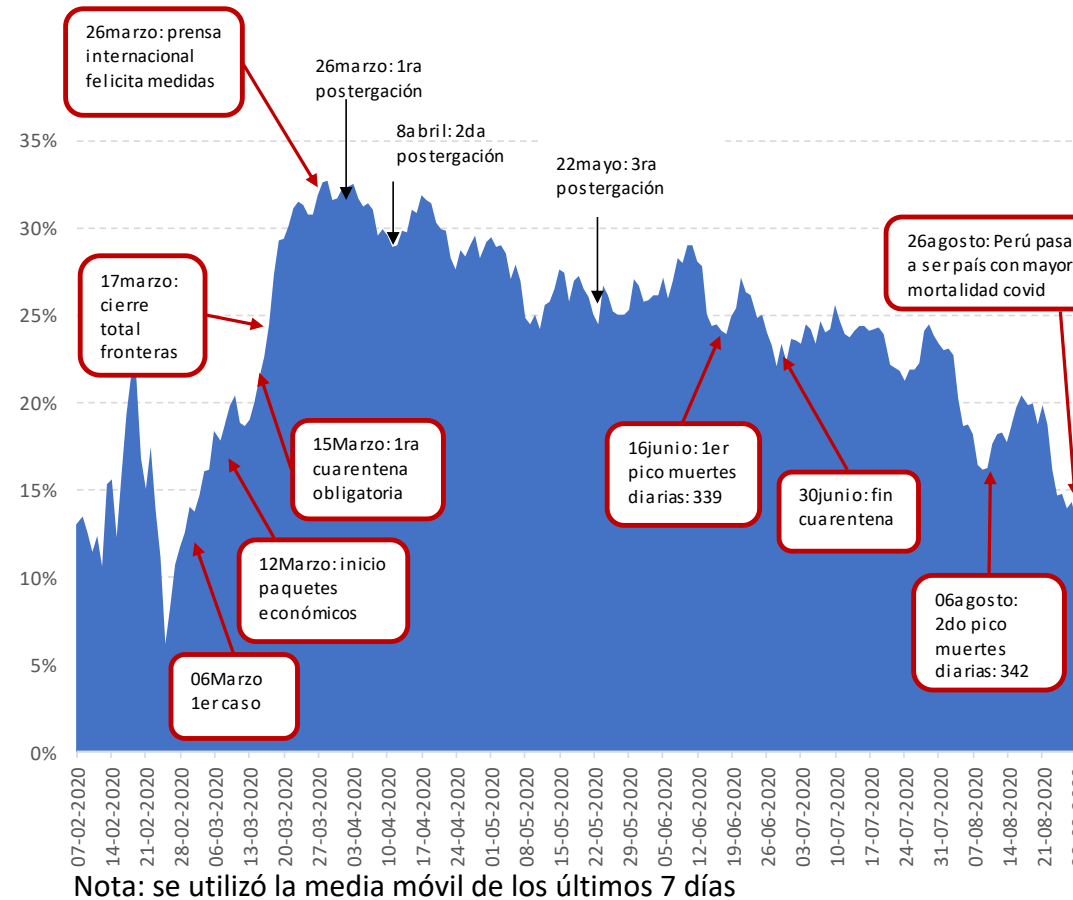


"¡Se apareció un arco iris al comenzar el toque de queda! Ojalá ese arco iris sea un signo que las cosas mejorarán..." #quedateencasa #cuarentena en San Juan de Miraflores

Tweet del 31 de marzo, Perú

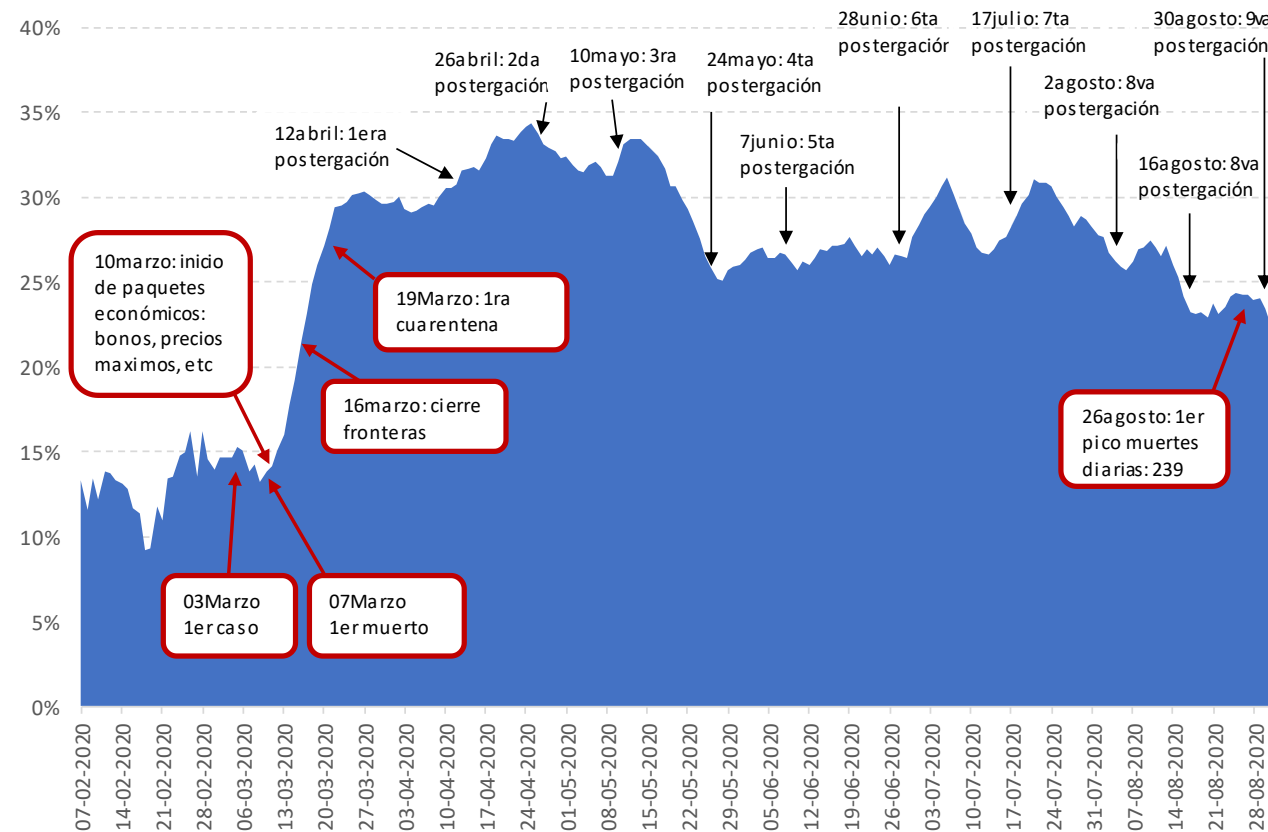
# Perú

**Gráfico 7.** Evolución de ratio de positividad en Perú, febrero-agosto de 2020



# Argentina

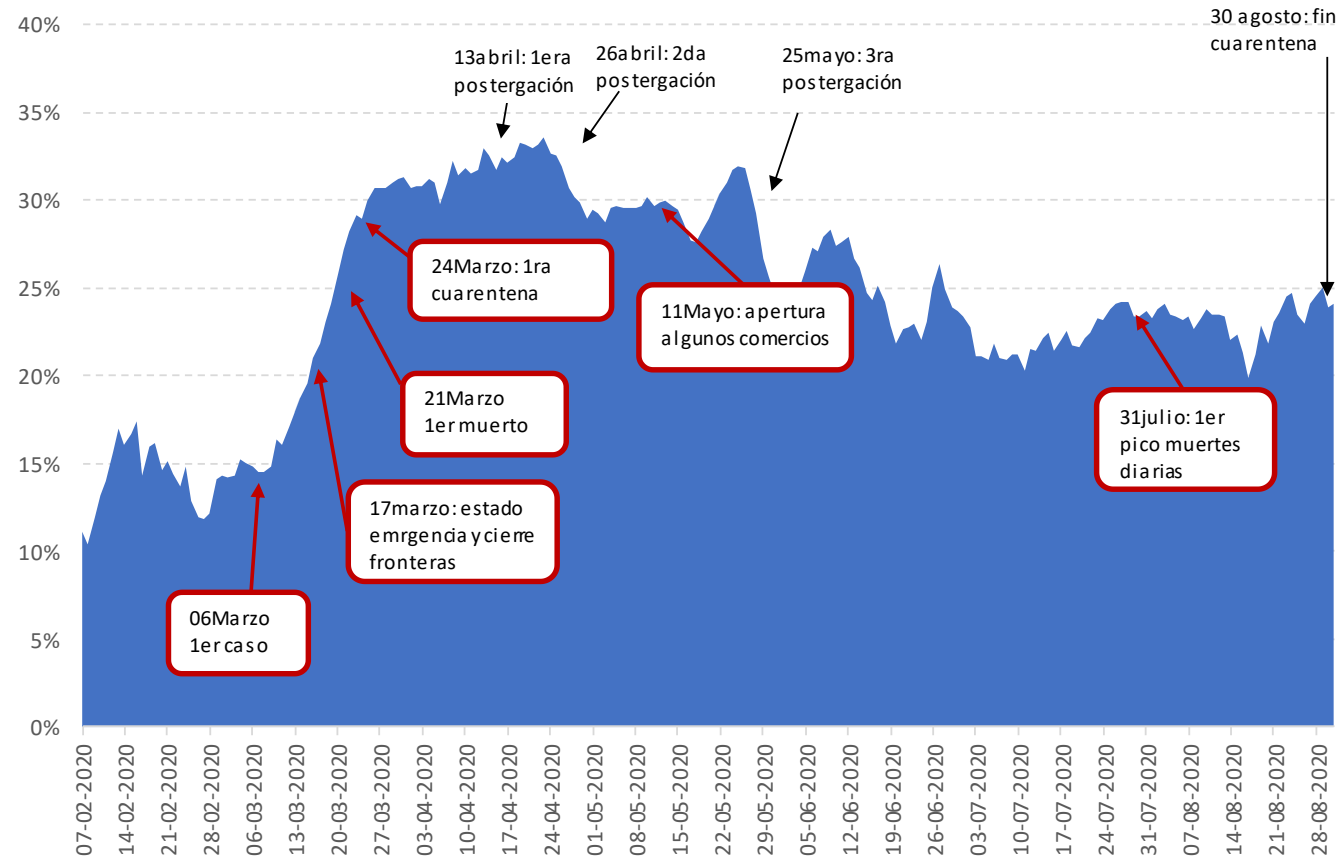
**Gráfico 8.** Evolución de ratio de positividad en Argentina, febrero-agosto de 2020



Nota: se utilizó la media móvil de los últimos 7 días

# Colombia

**Gráfico 9.** Evolución de ratio de positividad en Colombia, febrero-agosto de 2020

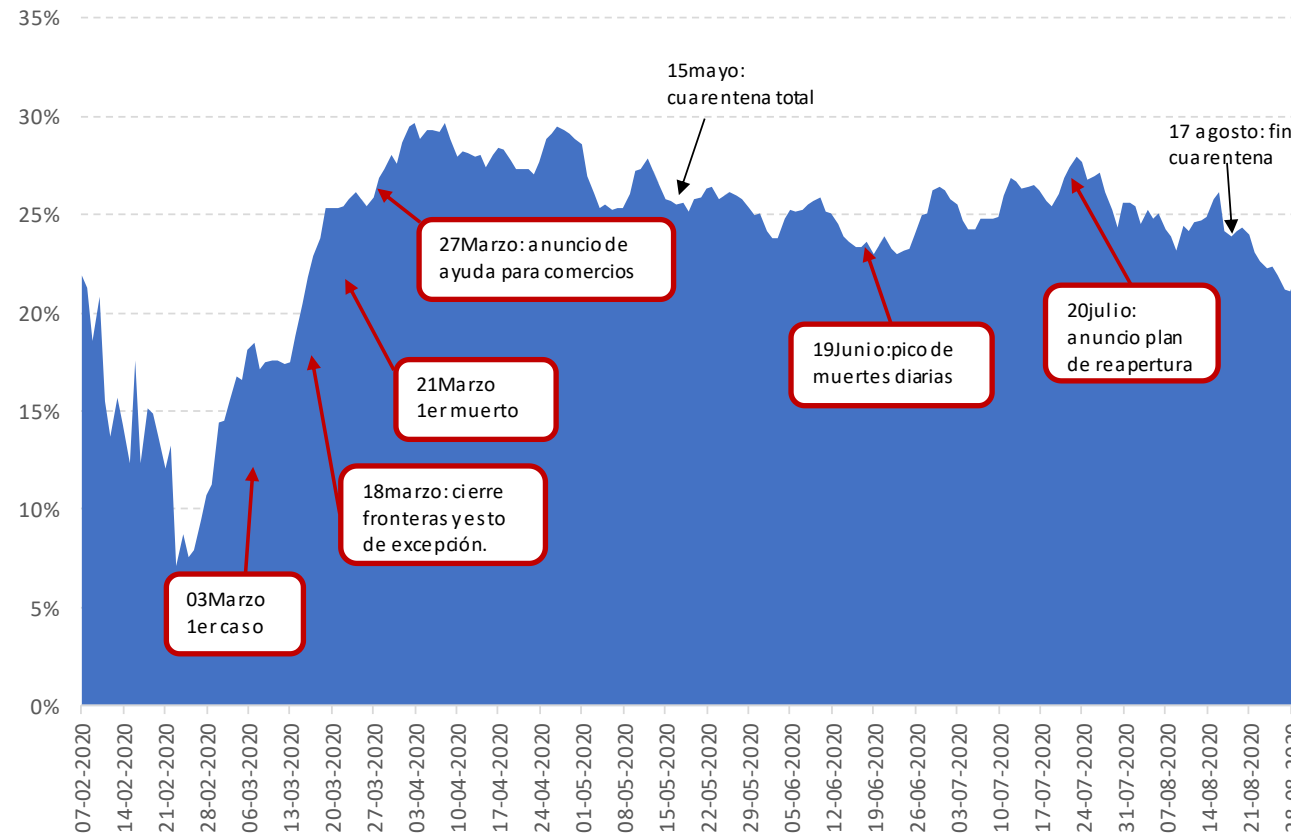


Nota: se utilizó la media móvil de los últimos 7 días



# Chile

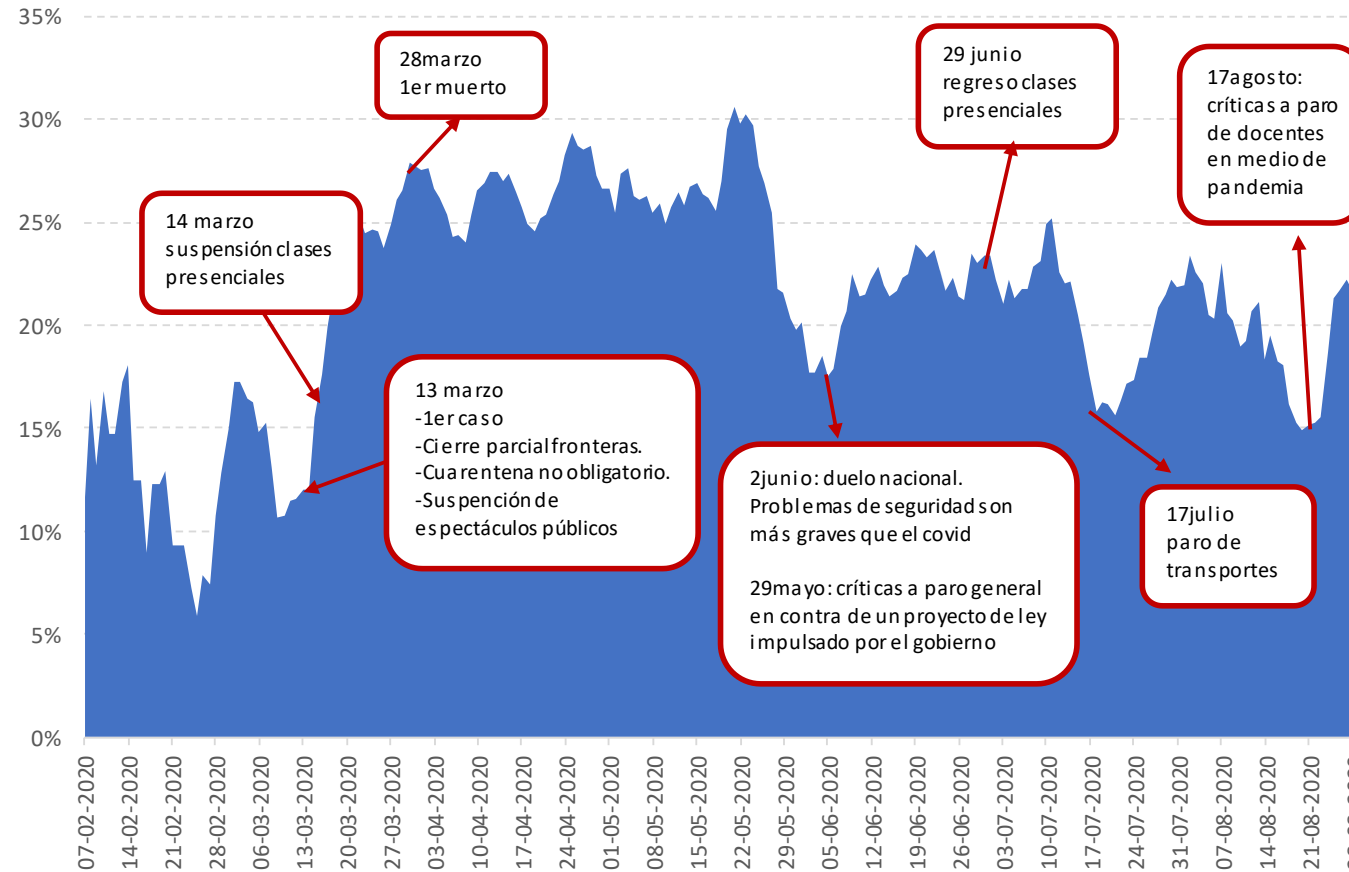
**Gráfico 11.** Evolución de ratio de positividad en Chile, febrero-agosto de 2020



Nota: se utilizó la media móvil de los últimos 7 días

# Uruguay

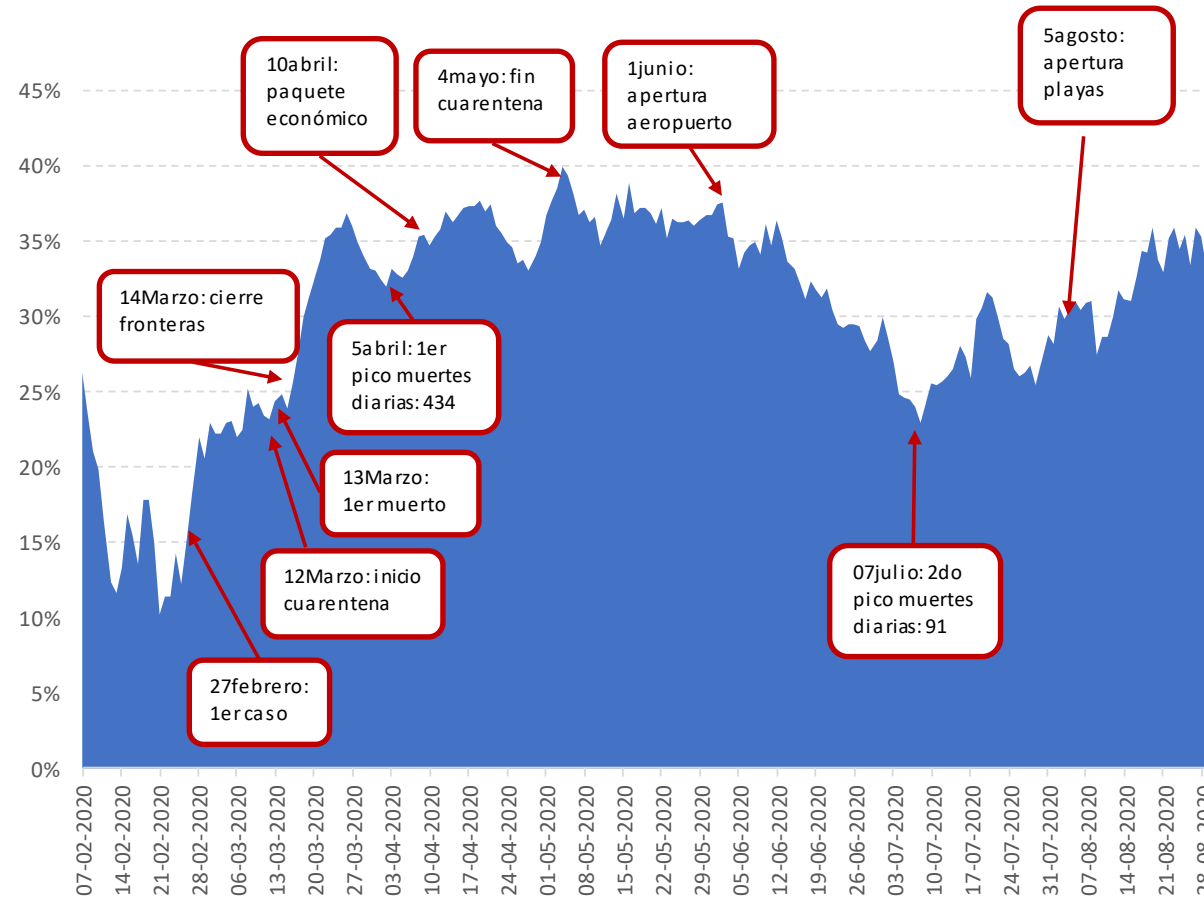
**Gráfico 12.** Evolución de ratio de positividad en Uruguay, febrero-agosto de 2020



Nota: se utilizó la media móvil de los últimos 7 días

# Ecuador

**Gráfico 10.** Evolución de ratio de positividad en Ecuador, febrero-agosto de 2020

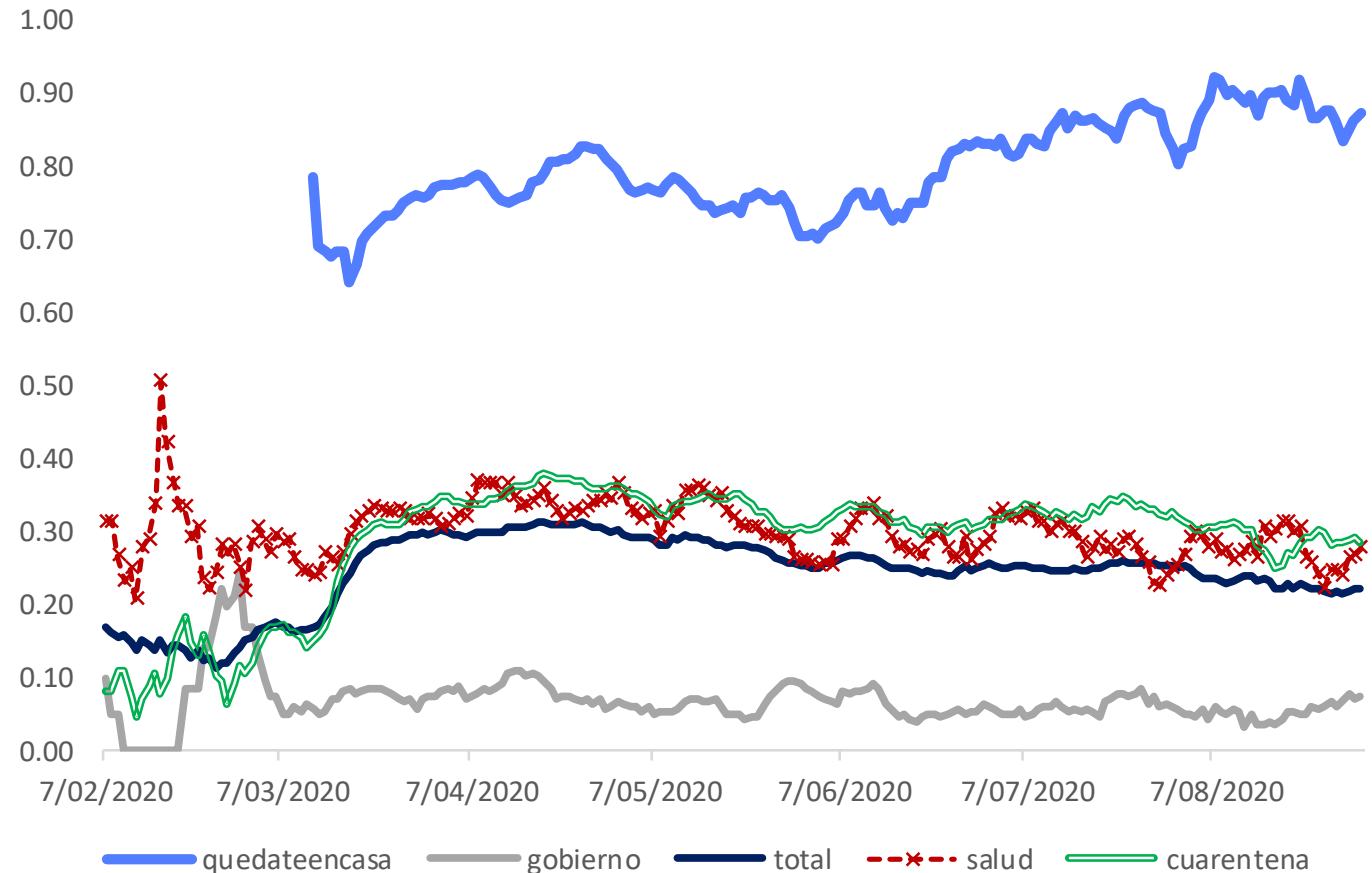


Nota: se utilizó la media móvil de los últimos 7 días

## Temas de interés

- Se analizan los siguientes temas: "Quédateencasa", "Gobierno", "Salud" y "Cuarentena"
- "Quédateencasa" tiene un porcentaje de positividad mucho mayor que "Cuarentena"
- "Gobierno" presentan ratios de positividad muy bajos (menor a 10%)
- Salud tiene una positividad similar al indicador de positividad promedio relacionado con términos de la COVID-19 (pandemia, coronavirus, entre otros)

**Gráfico 13.** Ratio de positividad de Tweets de términos seleccionados, febrero-



Nota: se utilizó la media móvil de los últimos 7 días

## Conclusiones

- Se encontró que todos los países mostraron valores menores a 40% en el periodo analizado; es decir los tweets relativos al coronavirus se asociaron a sentimientos negativos
- En promedio, Ecuador y Argentina son los países analizados que presentan mayores niveles de positividad a lo largo de los meses. Por otro lado, Perú y Uruguay mostraron los niveles de positividad más bajos.
- La población no está muy conforme con la actuación del Gobierno
- Los tweets con el hashtag Quédateencasa presentan un sentimiento de positividad mucho mayor que el de aquellos con la palabra “Cuarentena” lo que podría significar una oportunidad para que los gobiernos enfoquen mejor la manera de comunicación de las medidas ante posibles nuevos rebrotes que vuelvan a obligar a la población a permanecer en sus casas.

## Conclusiones

- Se encontraron las siguientes limitaciones para la elaboración del presente trabajo:
  - ✓ El etiquetado automático con emojis puede llevar a contradicciones
  - ✓ No se consideraron las opiniones neutrales ni sentimientos más complejos
  - ✓ La geolocalización no es exacta
  - ✓ Los países de Sudamérica no cuentan con una penetración alta en el uso de Twitter (inferencia limitada)

## Bibliografía

[Barkur, G.](#), [Vibha](#), y Kamath, [G.](#) (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry*. (51).

<https://doi.org/10.1016/j.ajp.2020.102089>

Dubey, A. D. (2020). *Twitter Sentiment Analysis during COVID-19 Outbreak*.

<http://dx.doi.org/10.2139/ssrn.3572023>

Köksal A. (2020). *BERT Sentiment Analysis Turkish*

<https://github.com/akoksal/BERT-Sentiment-Analysis-Turkish>

Korkut, U., Foley, J. y Ozduzen, O. (2020). The Digital Publics of #Schengen and #Eurozone During the Coronavirus Crisis. *Respond*. (3).

<https://drive.google.com/file/d/1f8uokBgrptSgGwNpQe-beyDSOJpQ1wRg/view>

# Bibliografía

Liu, B., (2012). *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers  
<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>

Samuel, J., Nawaz, G., Rahman, M., Esawi y E., Samuel, Y. (2020). *COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification*.  
<https://doi.org/10.3390/info11060314>

[Sharma](#), K., [Seo](#), S., [Meng](#), C., [Rambhatla](#), S. y [Liu](#), Y. (2020). *COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations*.  
<https://arxiv.org/abs/2003.12309>

Sobrino, J.C. (2018). *Análisis de Sentimientos en Twitter* [Tesis de maestría, Universidad Oberta de Catalunya].  
<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81435/6/jsobrinostFMo618memoria.pdf>



## Bibliografía

Storjohann, P. (2005), *Corpus-driven vs. corpus-based approach to the study of relational patterns*

[https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/5006/file/Storjohann\\_Corpus\\_driven\\_vs\\_corpus\\_based\\_approach\\_to\\_the\\_study\\_of\\_relational\\_patterns\\_2005.pdf](https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/5006/file/Storjohann_Corpus_driven_vs_corpus_based_approach_to_the_study_of_relational_patterns_2005.pdf)

Zhang, M., Ng, J., (2020), *Twitter Sentiment Analysis: What does Social Media tell us about coronavirus concerns in the UK?*

<https://www.actuaries.org.uk/system/files/field/document/Twitter%20Sentiment%20Analysis.pdf>



## COVID-19 EN REDES SOCIALES: ANÁLISIS DE SENTIMIENTO EN SUDAMÉRICA, 2020 \*

Pilar Villena Guzmán  
pilarvillena@edu.uah.es

Jillie Chang Kcomt  
jillie.chang@edu.uah.es

**Universidad Alcalá de Henares (UAH) Madrid**  
**Máster en Business Intelligence y Data Science**  
Asesor: Dr. Lino González García

\* Código y base disponibles en <https://github.com/TFMChangVillena/AnalysisSentimentCovidSudamerica>

# Anexo

Evolución de ratio de positividad en Colombia, febrero-agosto de 2020

