



# Linda Cabrera Orellana

Debate en Twitter sobre el encarcelamiento  
de PABLO HASÉL, usando Procesamiento  
de Lenguaje Natural



**WOMEN IN DATA SCIENCE**  
ECUADOR

**WiDS Quito**

# AGENDA

Introducción

Análisis de datos

Procesamiento de lenguaje natural

Tokenización y lematización

Análisis de sentimientos

# Datos informativos de Twitter

Fuente: <https://kinsta.com/es/blog/estadisticas-twitter/>



15% Bots



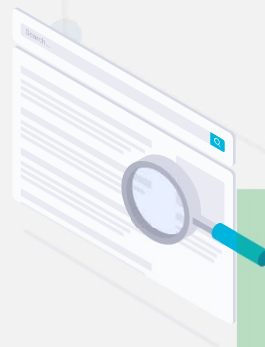
Usuarios

46% de los usuarios activos de Twitter utilizan el sitio a diario

Usuarios

1.3 billones de cuentas de Twitter

335 millones de usuarios activos al mes



Tweets

456,000 Tweets por minuto  
656 millones de Tweets al día

# Los datos de Twitter



- 👉 API de Twitter es abierta
- 👉 Límites en la descarga de tweets en tiempo y cantidad
- 👉 280 caracteres por tweet
- 👉 Información geográfica deficiente

# Análisis de datos



👉 library(rtweet)

👉 @PabloHasel OR \"Pablo Hasel\"

👉 336,148 tweets desde el 19 al 28 de febrero de 2021

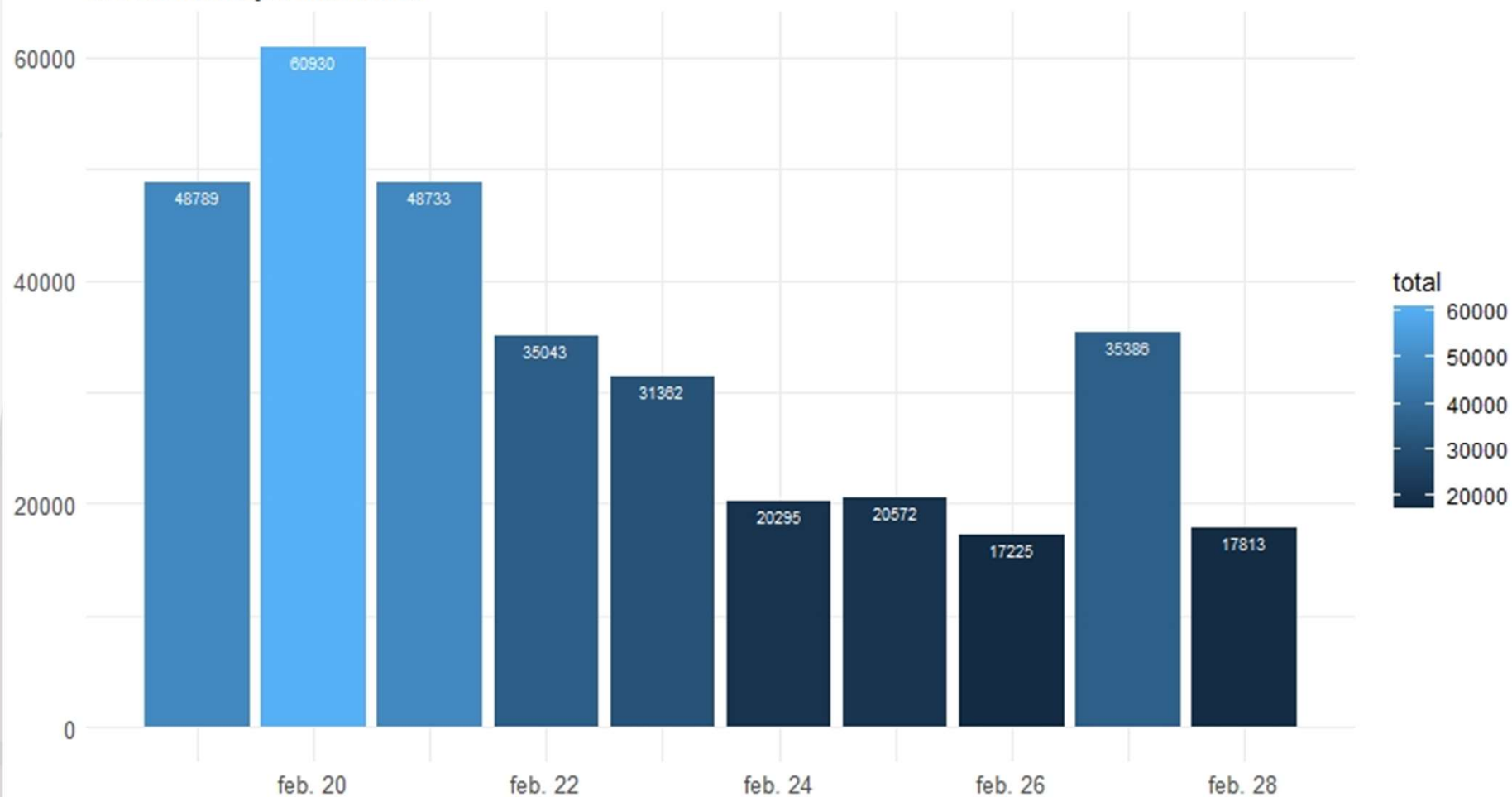
👉 91 variables

👉 Algunos componentes son:

- "user\_id", "status\_id", "created\_at"
- "text", "is\_retweet", "favorite\_count"
- "retweet\_count", "hashtags", "mentions\_user\_id"
- "geo\_coords", "followers\_count"

**Figura 1: Tweets posteados por días desde el 19 al 28 de febrero del 2021**

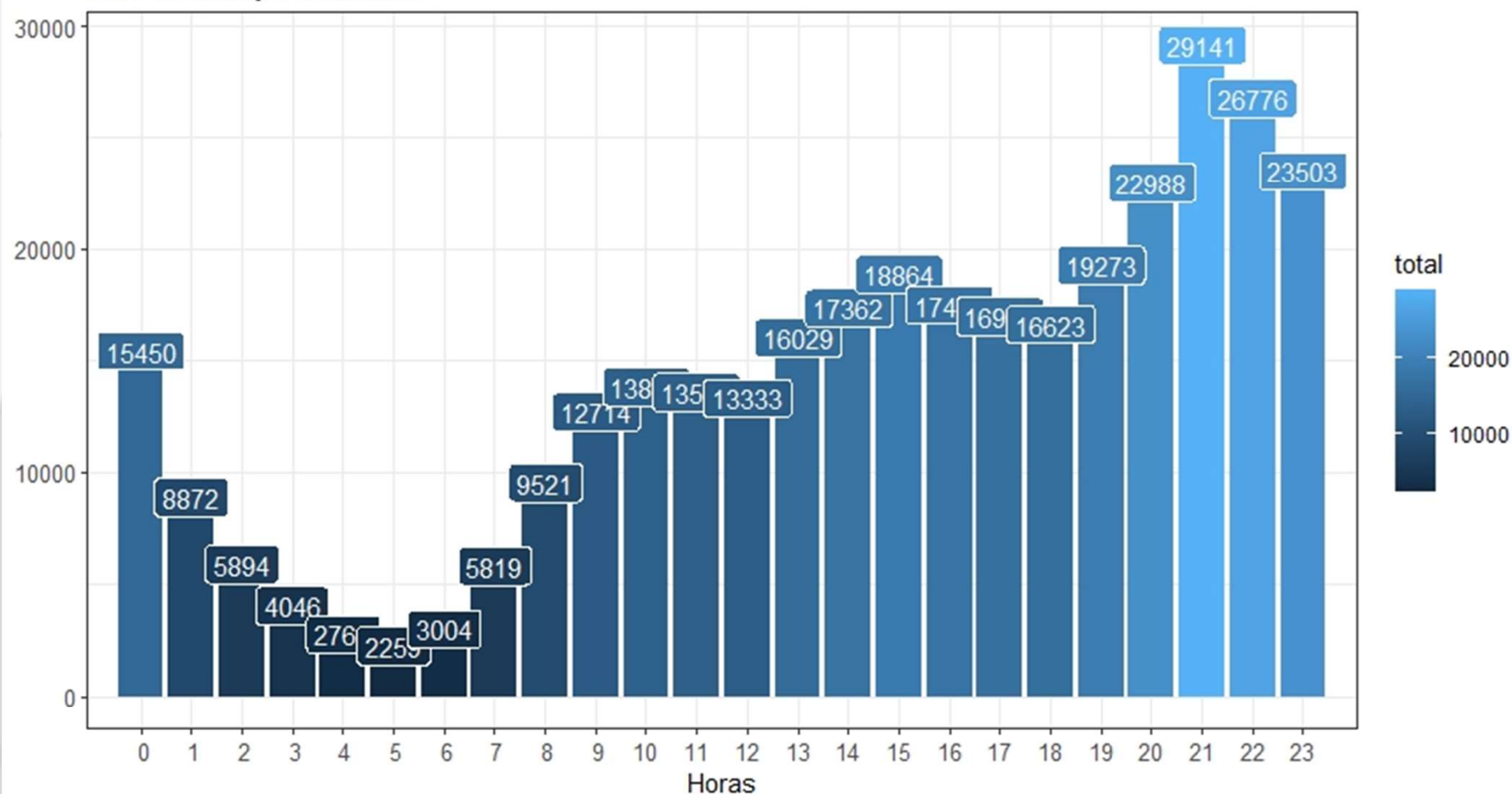
336148 mensajes analizados



Source: Data collected from Twitter's REST API via rtweet

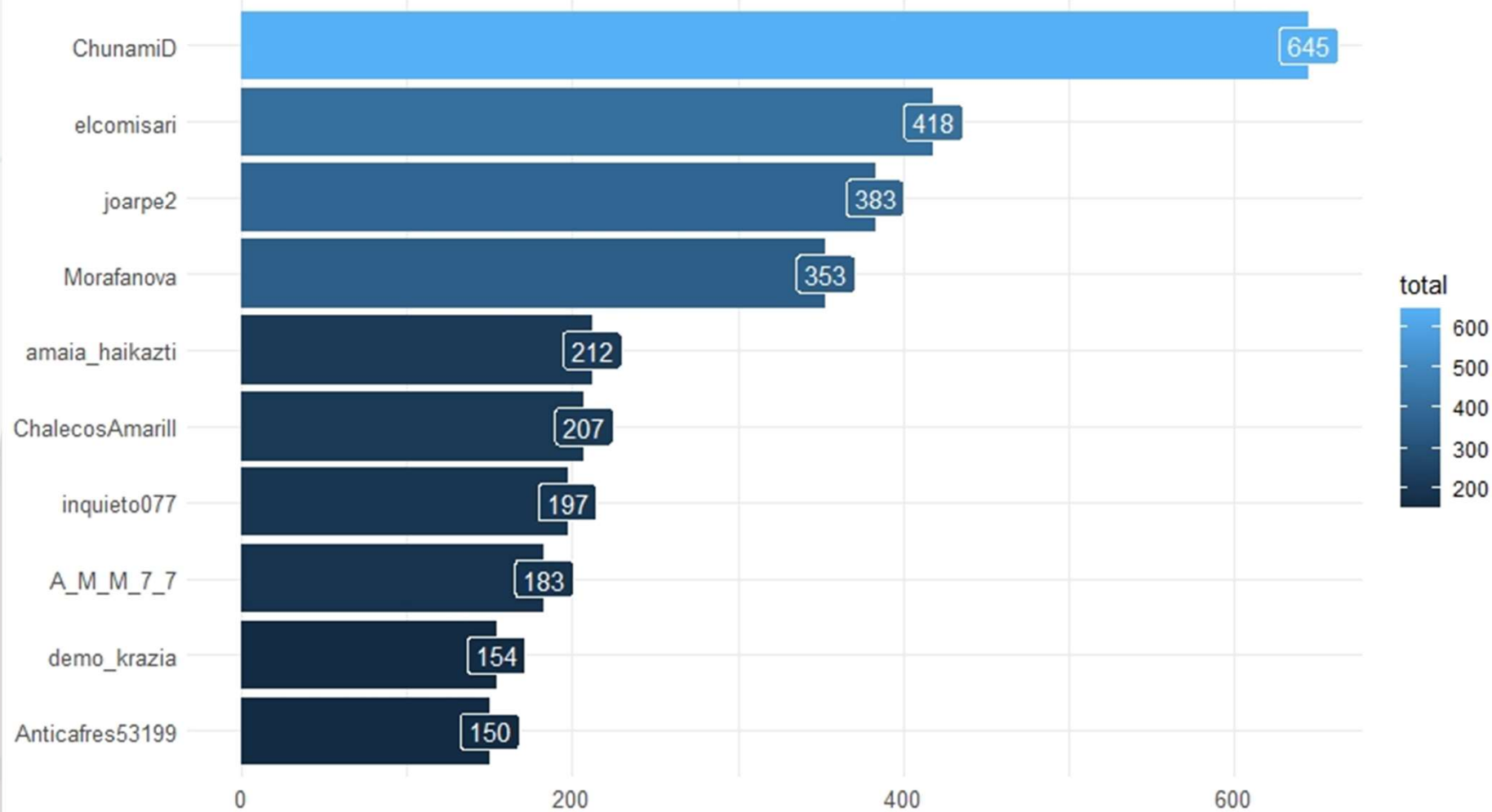
**Figura 2: Tweets por Hora de posteo (Madrid/España)**

336148 mensajes analizados



Source: Data collected from Twitter's REST API via rtweet

**Figura 3: TOP 10 - Cuentas que más postean sobre Pablo Hasél**

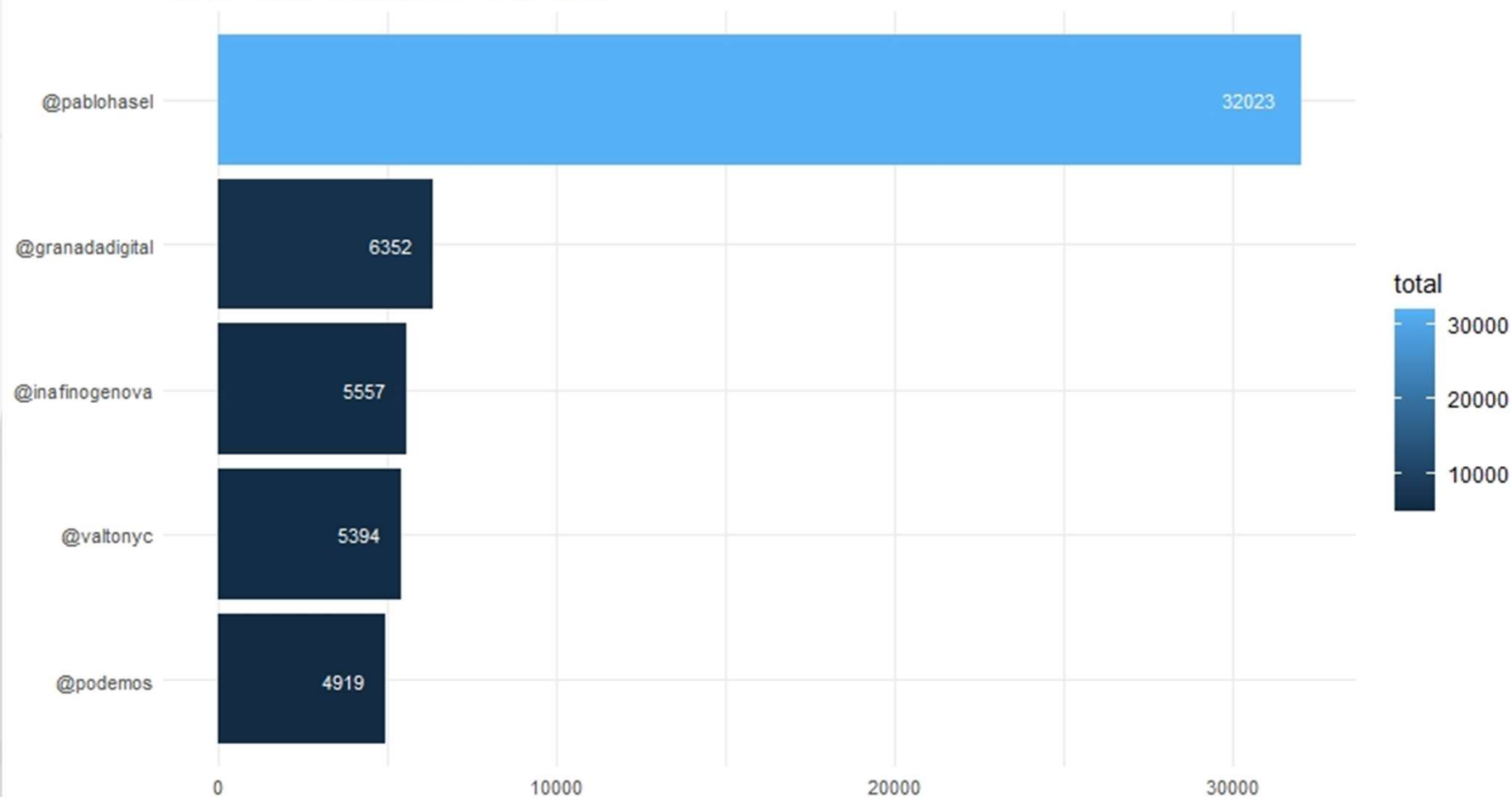


Source: Data collected from Twitter's REST API via rtweet



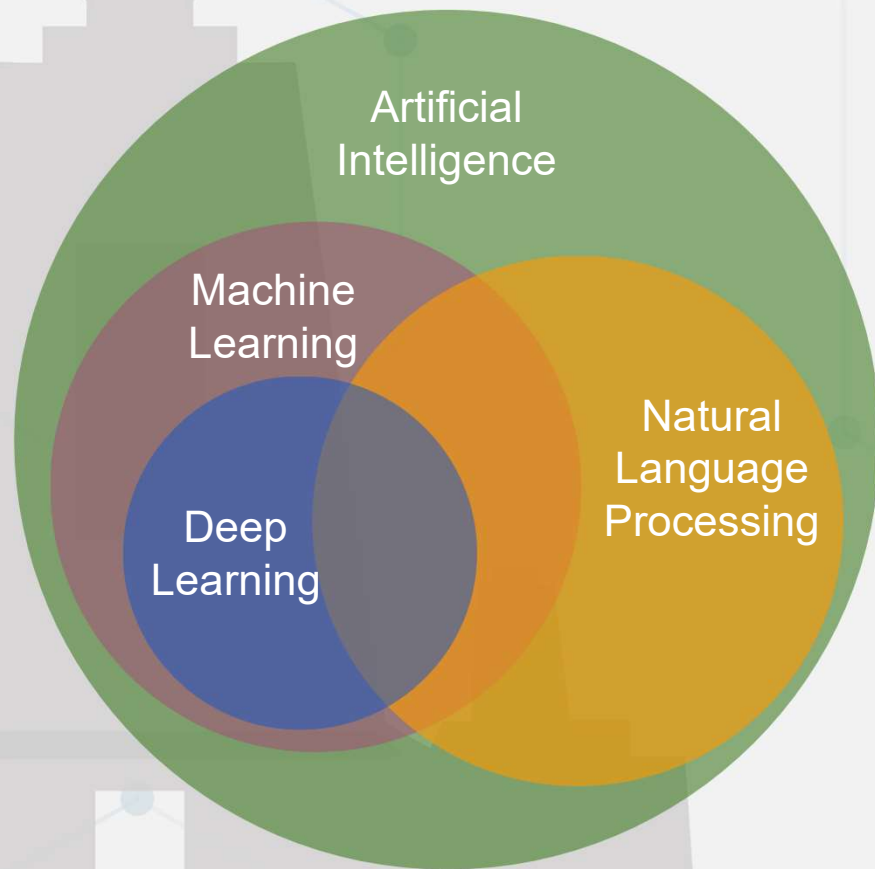
**Figura 4: TOP 5 - Cuentas más mencionadas**

En los tweets hablando de Pablo Hasel



Source: Data collected from Twitter's REST API via rtweet

# Procesamiento de lenguaje natural PLN



Según Sarkar (2019):

*“El PLN se define como un campo especializado de las ciencias de la computación, la ingeniería y la inteligencia artificial enraizado en la lingüística computacional. Este está principalmente enfocado en el diseño y construcción de aplicaciones y sistemas que permitan la interacción entre máquinas y lenguajes naturales utilizados por el ser humano... Las técnicas de PLN permiten que las computadoras procesen y entiendan el lenguaje natural humano y lo utilicen para dar resultados útiles.” (traducido del inglés).*

# Tokenización



## Tweet original

"Echenique dice que está muy tranquilo porque las denuncias como siempre acabarán en nada. Será porque con las condenas que acumula considera que ha cubierto ya el cupo. <https://t.co/QbFrkaB0JH>"

## Tweet tokenizado

"echenique" "dice" "que" "esta" "muy" "tranquilo" "porque" "las" "denuncias" "como" "siempre" "acabaran" "en" "nada" "sera" "porque" "con" "las" "condenas" "que" "acumula" "considera" "que" "ha" "cubierto" "ya" "el" "cupo" "https" "tco" "QbFrkaB0JH"

# Stopwords



## Tweet tokenizado

```
"echenique" "dice" "que" "esta" "muy" "tranquilo" "porque"  
"las" "denuncias" "como" "siempre" "acabaran" "en" "nada"  
"sera" "porque" "con" "las" "condenas" "que" "acumula"  
"considera" "que" "ha" "cubierto" "ya" "el" "cupo" "https"  
"tco" "QbFrkaB0JH"
```

## Tweet tokenizado sin stopwords

```
"echenique" "tranquilo" "denuncias" "acabaran" "condenas"  
"acumula" "considera" "cubierto" "cupo"
```

# Palabras analizadas



# Tweets analizados

*48,789*

# Palabras después de tokenización

*1'311,355*

# Palabras sin stopwords

*704,330*

# Lematización y POS tags



## Tweet tokenizado sin stopwords

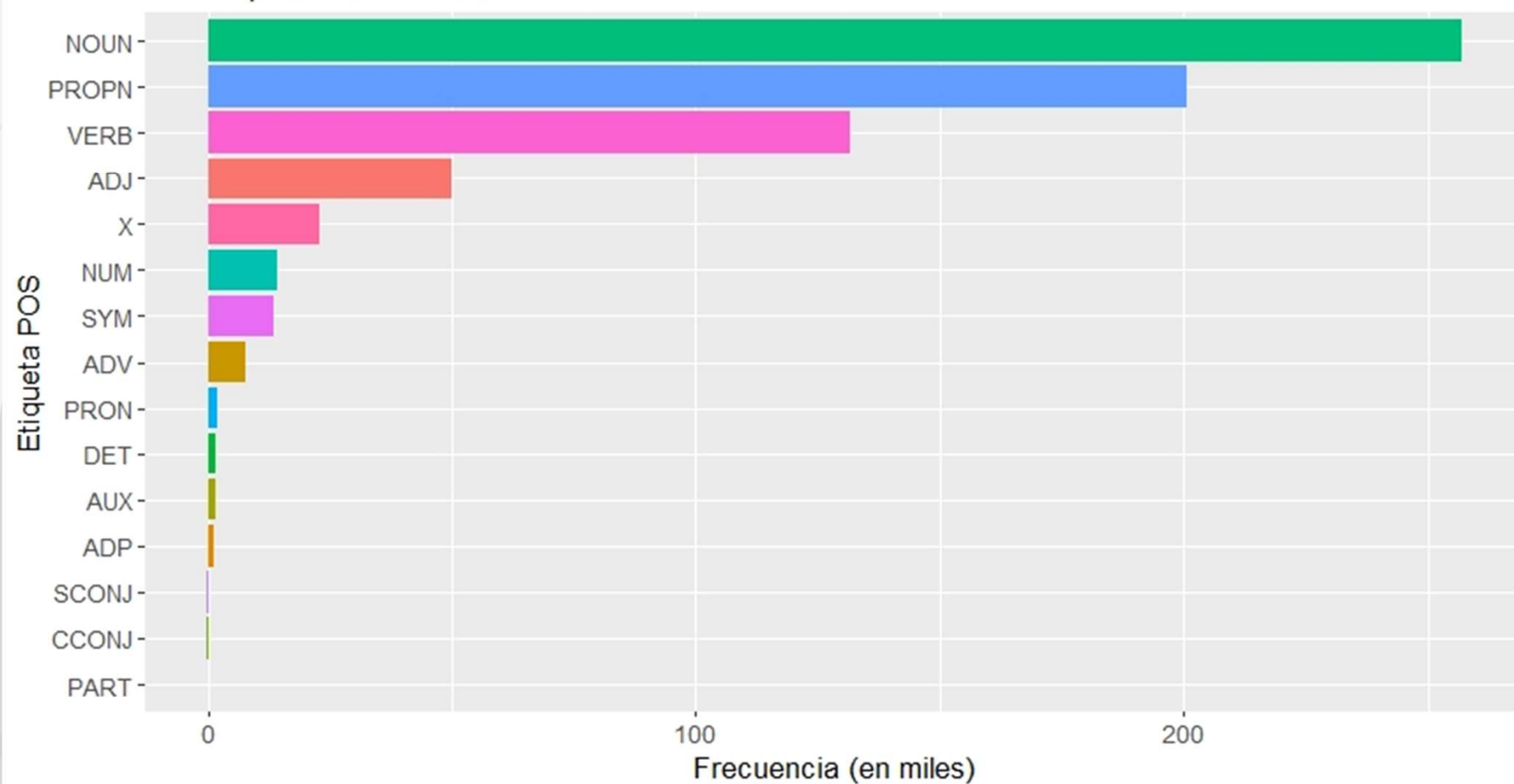
"echenique" "tranquilo" "denuncias" "acabaran" "condenas"  
"acumula" "considera" "cubierto" "cupo"

## Tweet lematizado y etiquetado gramatical

token	lemma	upos
Echenique	Echenique	PROPN
tranquilo	tranquilo	ADJ
denuncias	denuncia	NOUN
acabarán	acabar	VERB
condenas	condena	NOUN
acumula	acumular	VERB
considera	considerar	NOUN
cubierto	cubrir	VERB
cupo	cupo	NOUN

Figura 5: Distribución de etiquetas POST

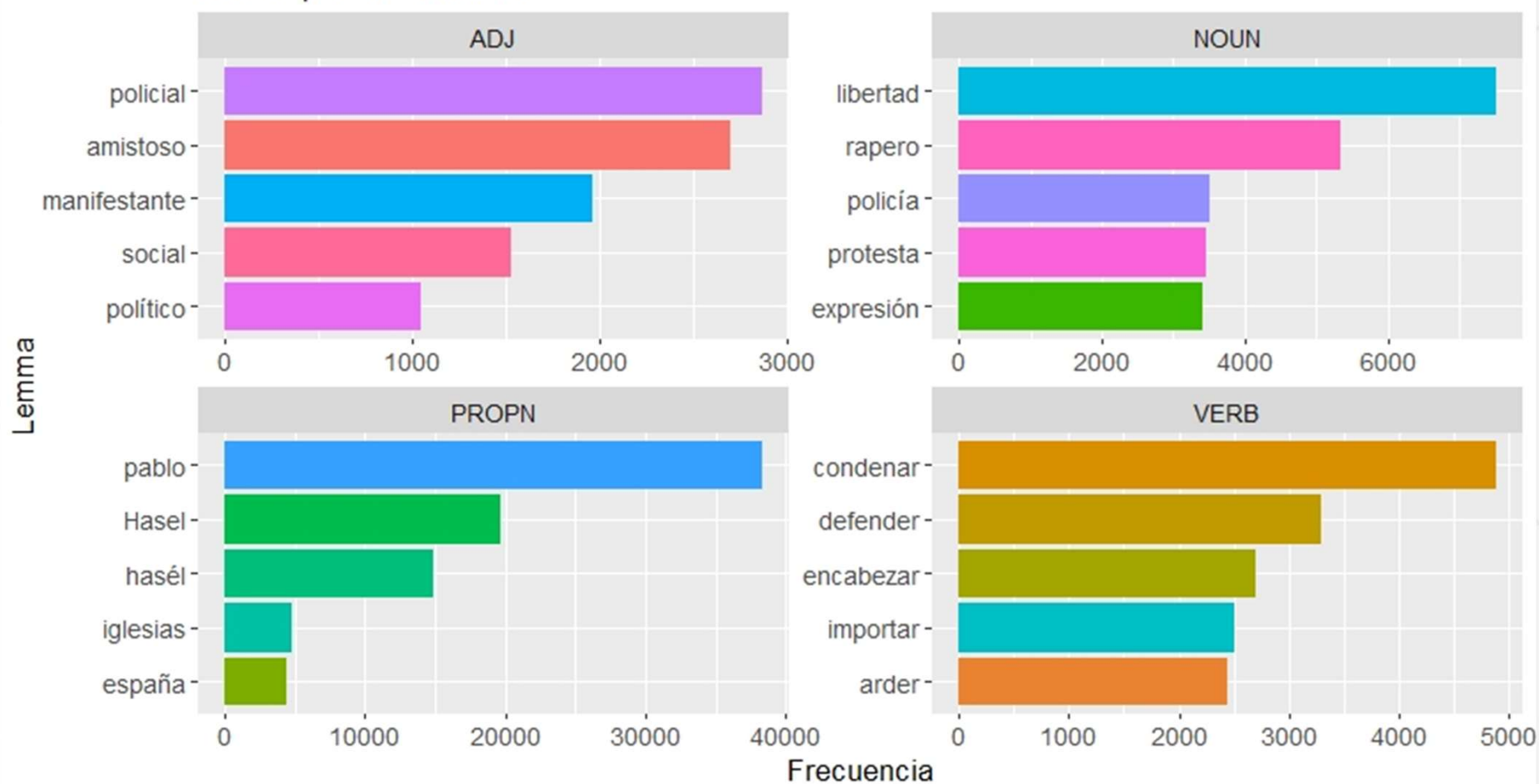
704330 palabras analizadas



Source: Data collected from Twitter's REST API via rtweet

Figura 6: Análisis por etiquetas POST

704330 palabras analizadas

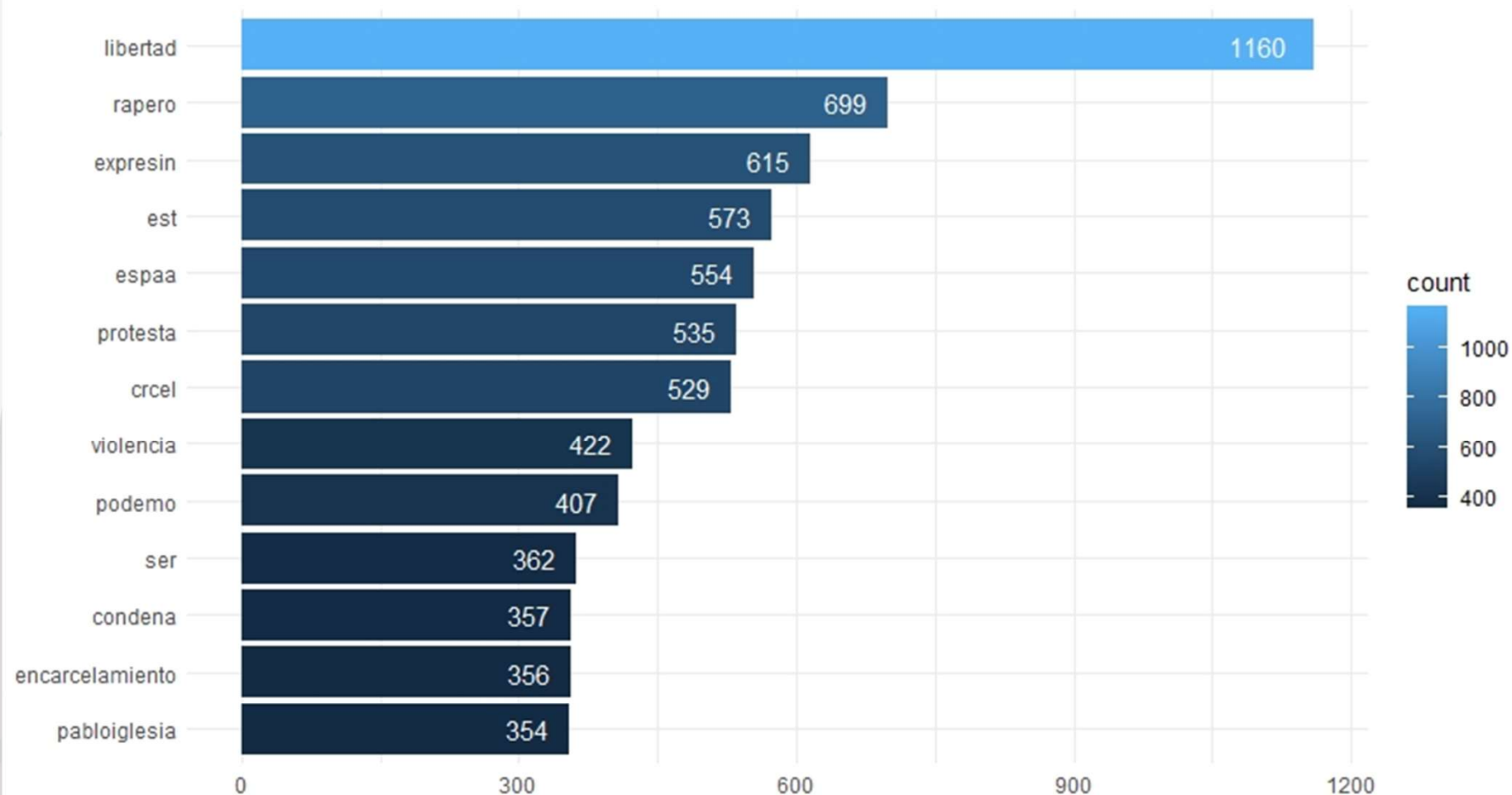


Source: Data collected from Twitter's REST API via rtweet



### Figura 7: Palabras más usadas

Conversaciones hablando de Pablo Hasel



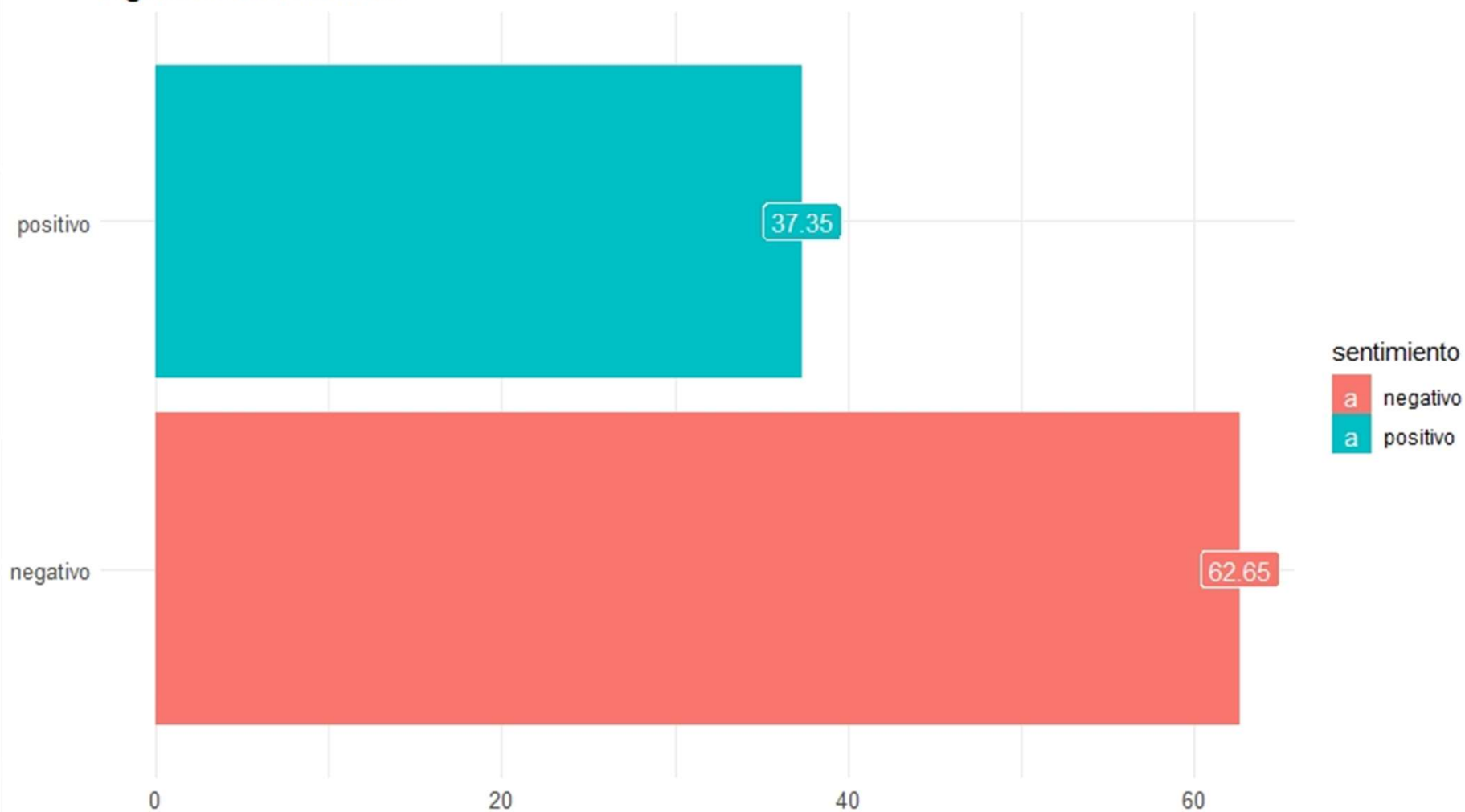
Source: Data collected from Twitter's REST API via rtweet

ochoabravo mestel donpinpon dan  
jordimasfont acvmatrix poltico disturbio odio aurantiacum lazarillatorm  
populacho crcel rapero pue ermanueamunik sargantania  
Igoritm barcelona testigo corona hoy monarqu cuenta delincuent hijo pena  
tavira carmatpor ytak jaum tambien claro favor violencia espaa manifestacin  
parec madrid creo mierda defend terrorismismo polica persona medio derecho  
ahora valtonyc hsel bien cosa tras van echeniqu directa juanma  
feder president mejor jcoscu izquierda oag detenidomundo gent delito prisin indepestil  
condena pabloiglesia enaltecimiento javierabaddaz rivadulla marcklein yucca  
sargantania chunamid mgadc inquieto acosta manifestacin tan sergio silvia lustr  
libertadpablohasel ciudad privado antifafr mismo morq verdad sabina dos christi  
condenado injuria mientras saez medio noch dos christi genovev gsuspesa vader  
fascista puta gar luquiti peibasqu  
juan ver ser ver mal vez aqu estn gobierno polica persona medio derecho  
hacer tuit snchez vdeo das part ley pas dice hace  
caso quin nadi manifestant  
corona hoy monarqu cuenta delincuent hijo pena  
pue ermanueamunik sargantania  
chunamid mgadc inquieto acosta manifestacin  
tan sergio silvia lustr  
libertadpablohasel ciudad privado antifafr mismo morq verdad sabina dos christi  
condenado injuria mientras saez medio noch dos christi genovev gsuspesa vader  
luquiti peibasqu  
enaltecimiento javierabaddaz rivadulla marcklein yucca

# Análisis de sentimientos

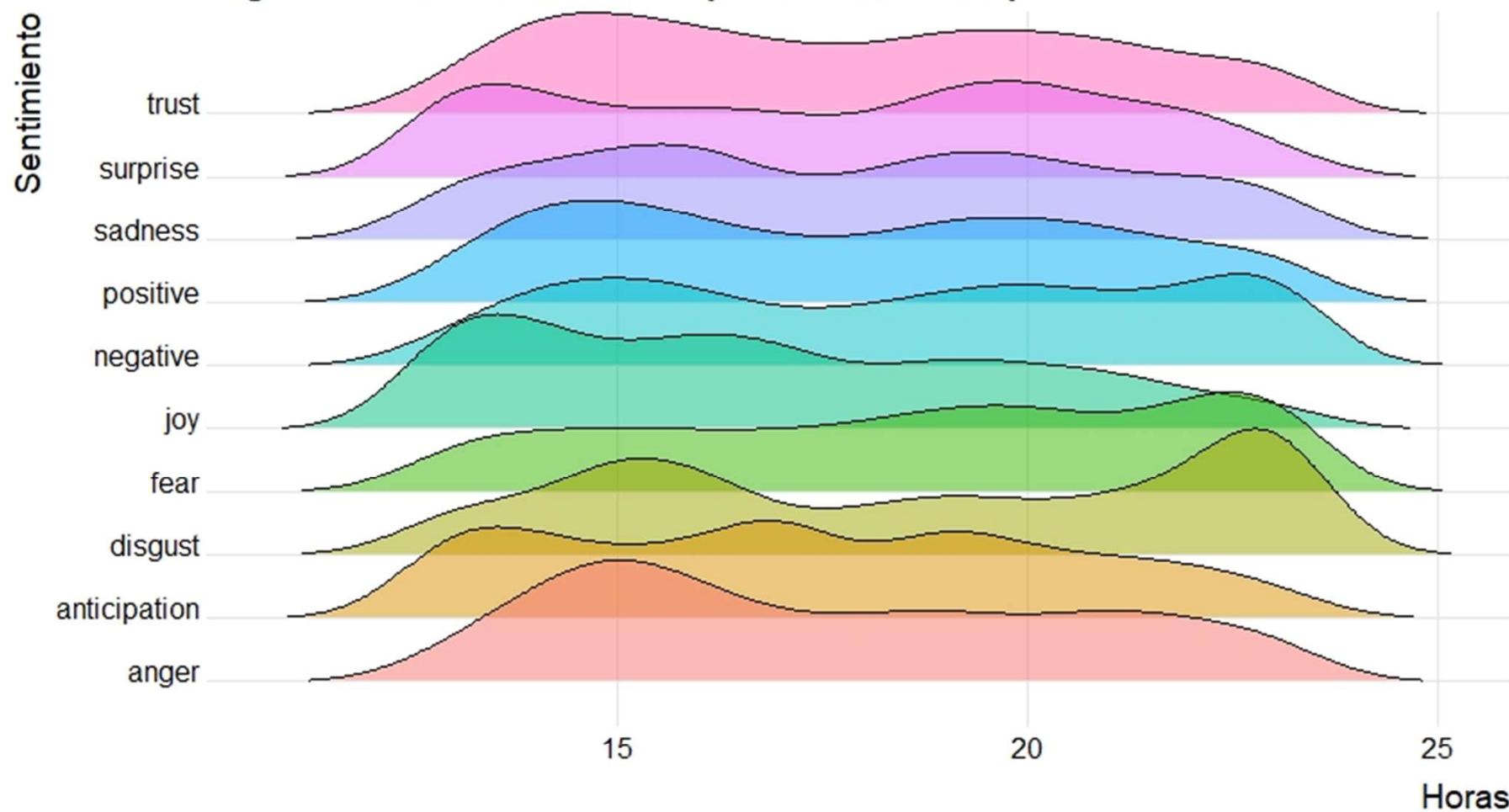


**Figura 9: Sentimientos**



Source: Data collected from Twitter's REST API via rtweet

**Figura 10: Sentimiento en los posteos en el tiempo**



Source: Data collected from Twitter's REST API via rtweet



The background is a light gray field featuring a network of thin, light blue lines that connect various points. Some points are solid blue circles, while others are solid black circles. On the left side, there is a large, semi-transparent, light purple geometric shape that resembles a stylized figure or a complex polygon. The word "CONCLUSIONES" is centered in the middle of the image in a green, sans-serif font.

# CONCLUSIONES



¡Gracias!

Linda Cabrera Orellana

UNIVERSIDAD TÉCNICA DE MACHALA



[ljcabrera@utmachala.edu.ec](mailto:ljcabrera@utmachala.edu.ec)



@linditacabrera