



BOT ACTIVITY ANALYSIS

**ARTE DE LA ANALÍTICA
ALEXIS-SOFÍA-GUSTAVO**

OBJETIVOS Y ANÁLISIS

Optimizar la visualización de la actividad de bots en Twitter por hora:

- Medible: Evaluar la claridad y la utilidad de las visualizaciones utilizando métricas de como los histogramas y correlaciones
- Alcanzable: Utilizar técnicas de visualización de datos más efectivas y comprensibles a través de gráficas en Python.
- Relevante: Comprender los patrones temporales de actividad de los bots es esencial para identificar posibles campañas coordinadas o actividades maliciosas.
- Con plazo: Implementar el análisis y evaluar el resultado en el plazo de la semana TEC

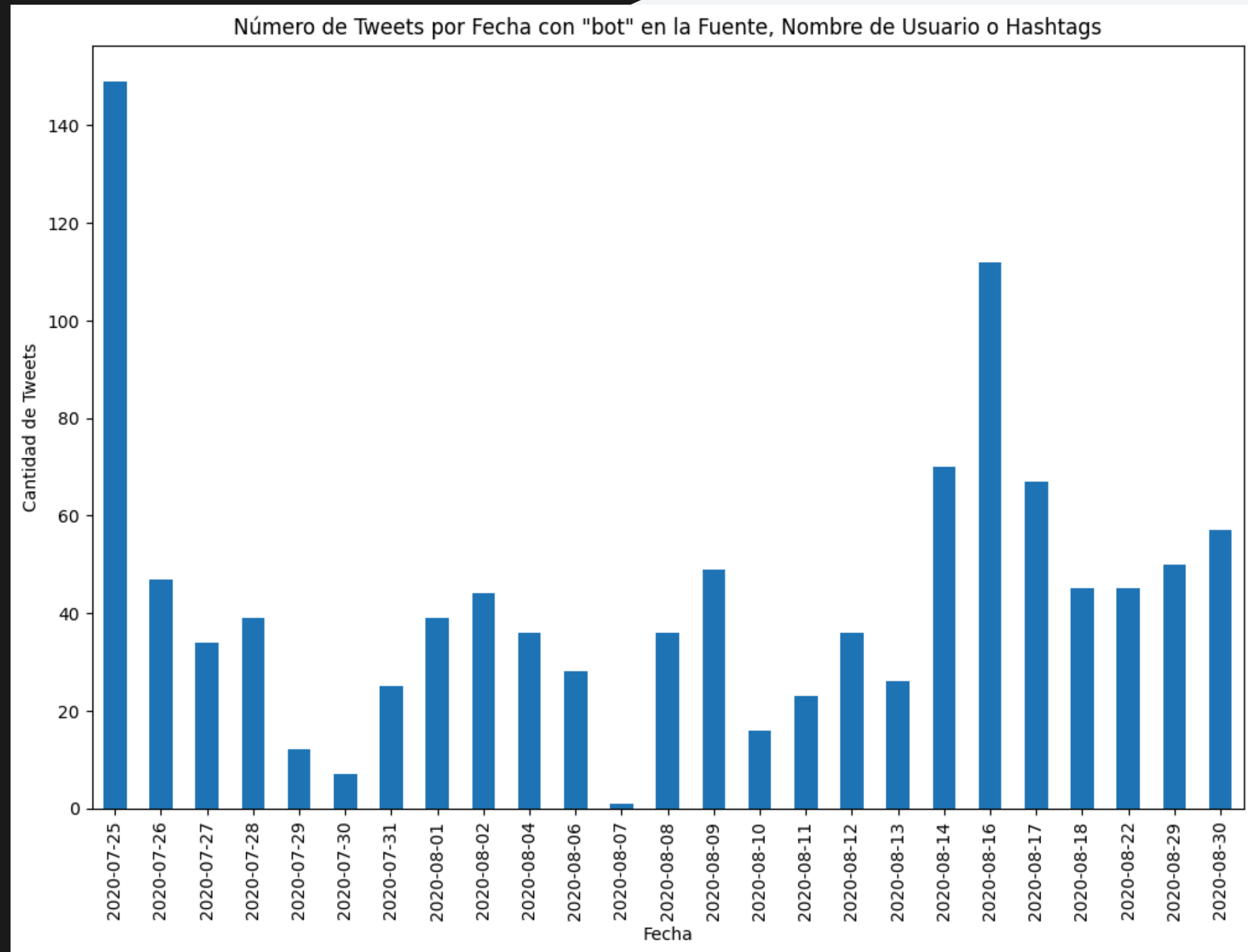
Incrementar la eficacia de la detección de bots en los datos de Twitter:

- Medible: Medir la mejora en la precisión de la detección de bots.
- Alcanzable: Utilizar algoritmos de detección de bots más avanzados o ajustar los criterios de filtrado para mejorar la precisión sin aumentar significativamente los recursos requeridos.
- Relevante: La detección precisa de bots es crucial para analizar la autenticidad y la influencia de las conversaciones en Twitter.
- Con plazo: Implementar el análisis y evaluar el resultado en el plazo de la semana TEC

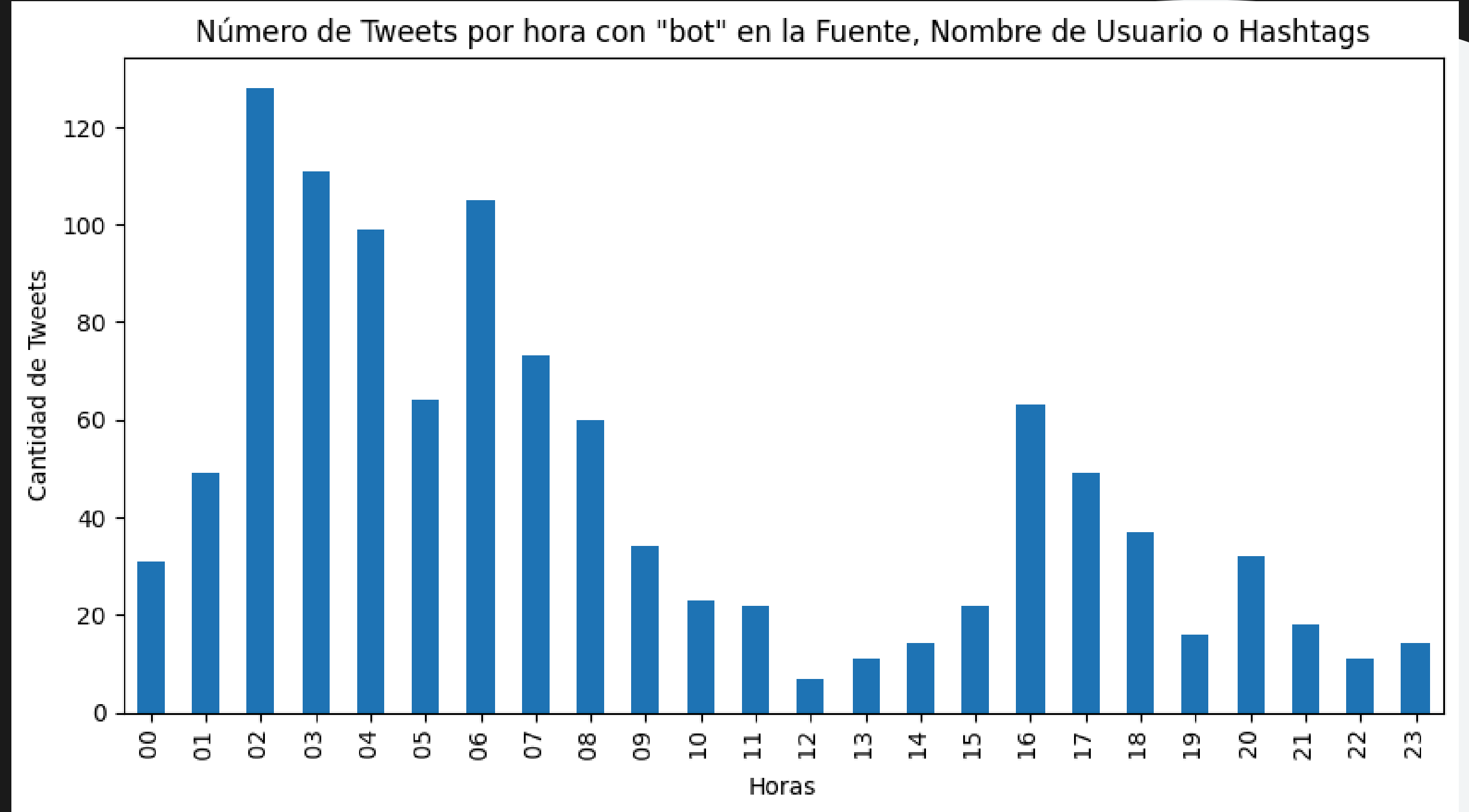
Mejorar la identificación de bots influyentes en Twitter:

- Medible: Evaluar la precisión de la identificación de bots influyentes mediante comparaciones. Mediante gráficas de cajas y bigotes, mapas de calor de correlación y clusters
- Alcanzable: Investigar y aplicar métodos más sofisticados de detección de bots influyentes, considerando tanto el número de seguidores como la actividad y la interacción.
- Relevante: Identificar bots influyentes es importante para comprender cómo se difunden y amplifican ciertos mensajes en Twitter.
- Con plazo: Implementar el análisis y evaluar el resultado en el plazo de la semana TEC

Este gráfico muestra la variación diaria en el número de tweets cuya fuente, nombre de usuario o hashtag diga "bot" entre el 25 de julio y el 30 de agosto de 2020, destacando un pico significativo el 13 de agosto Y un valor mínimo el 07 del mismo mes.

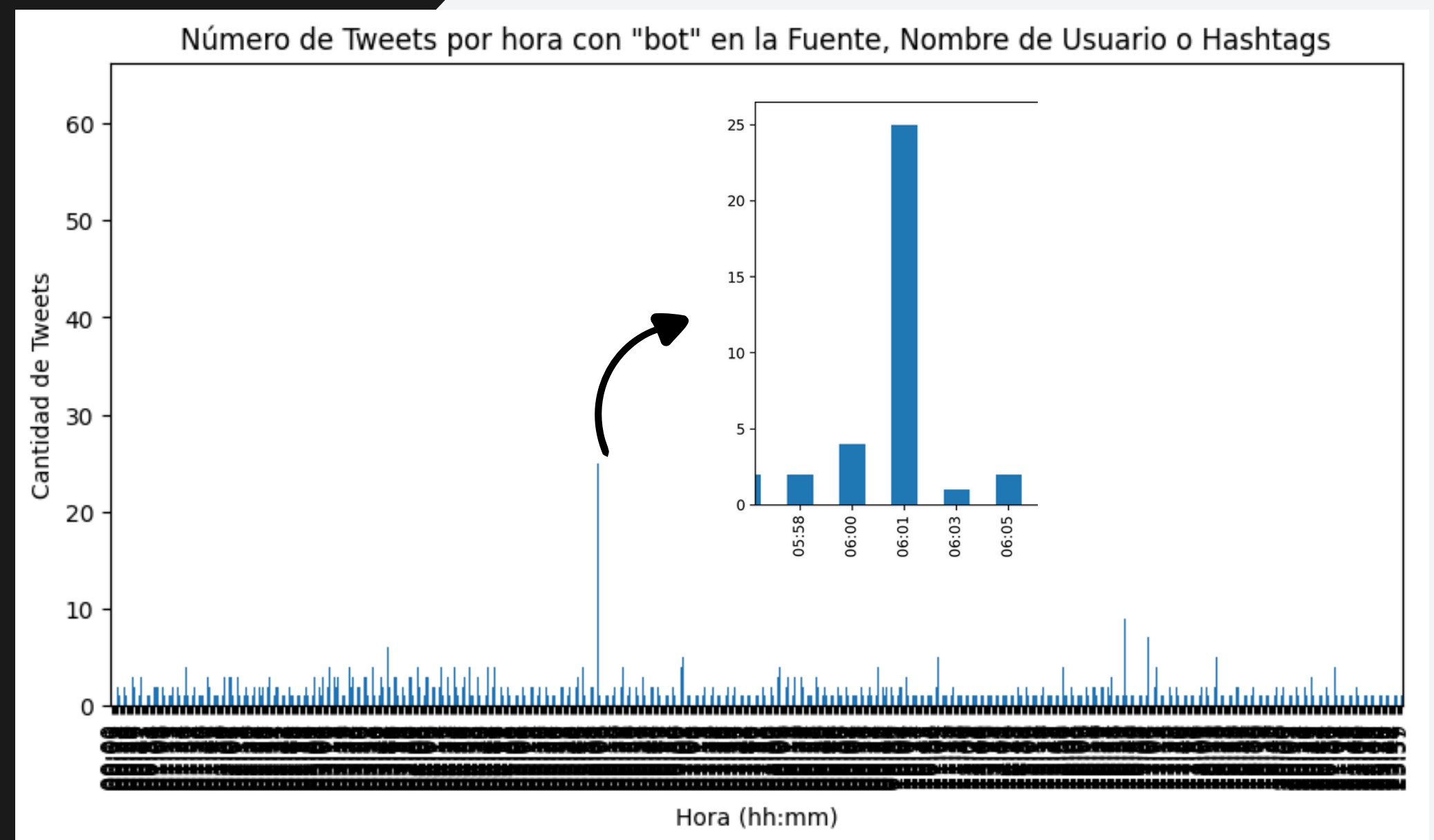


Este gráfico de barras representa el número de tweets que incluyen "bot" por hora. Se observa una actividad elevada entre las horas de la madrugada (00:00 a 04:00), con un segundo pico menos pronunciado en las primeras horas de la tarde (15:00 a 17:00).

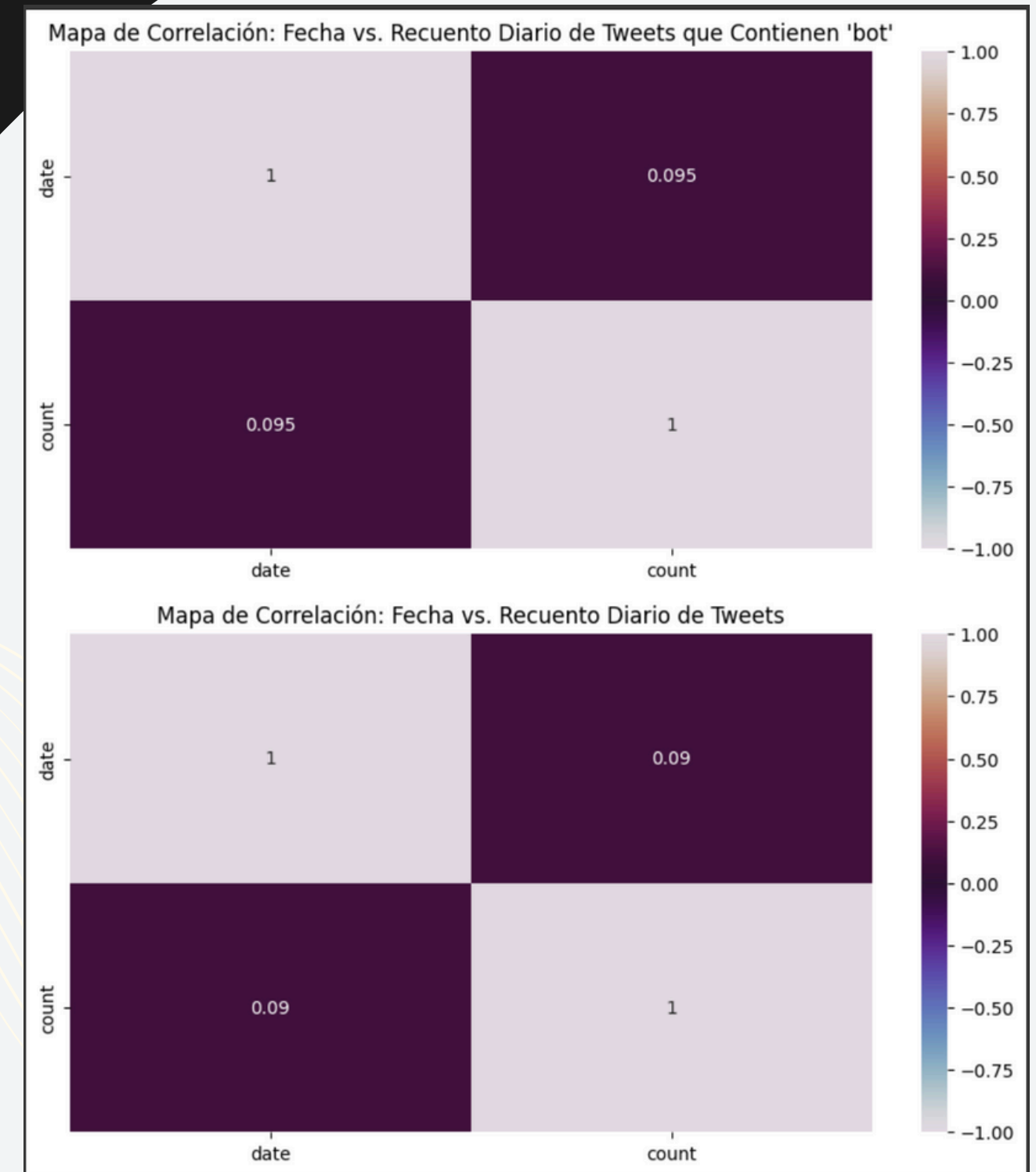


Este gráfico muestra la distribución de tweets por hora que mencionan "bot" en sus hashtags, en el nombre de usuario o en la fuente, registrando cada minuto del día.

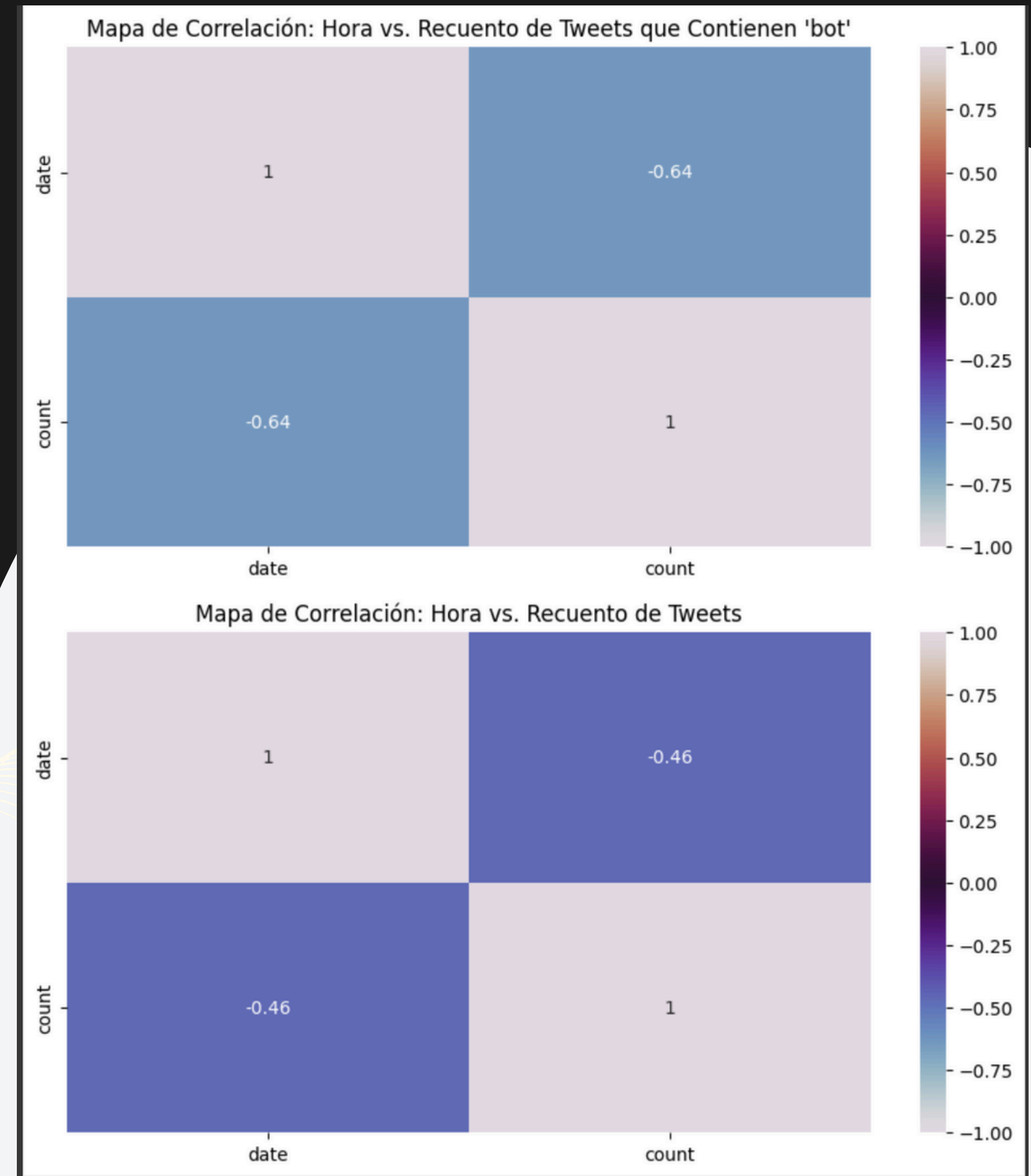
Destaca un pico prominente a las 6:01, con la mayoría de los otros momentos mostrando muy poca actividad relacionada con bots



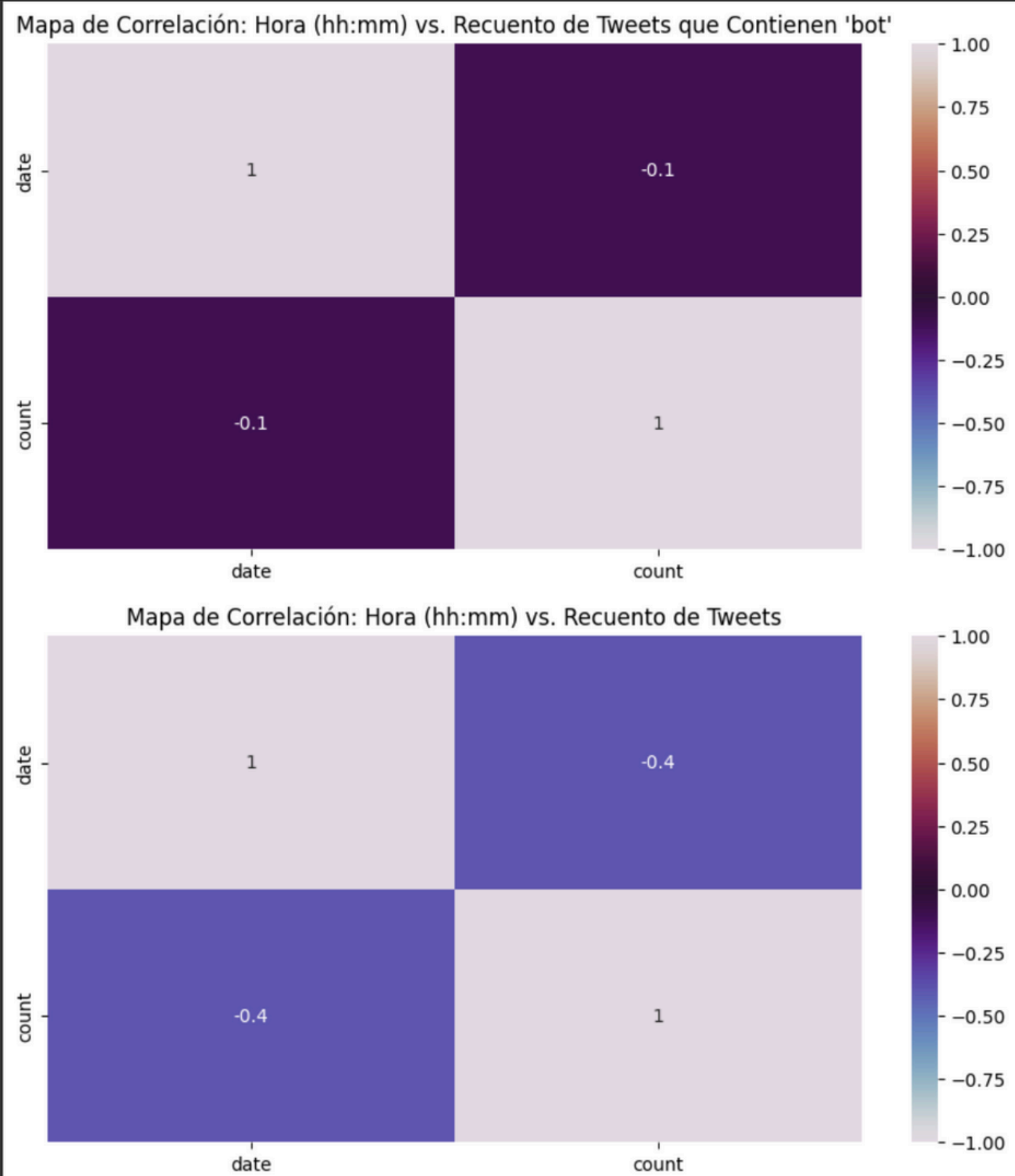
Este mapa de correlación muestra la relación entre las fechas y el recuento diario de tweets que contienen "bot". El valor 0.095 entre "fecha" y "recuento" indica una correlación muy baja, sugiriendo que no hay una relación lineal fuerte entre la fecha y la cantidad de tweets relacionados a "bot"



Este mapa de correlación examina la relación entre la hora del día y el recuento de tweets que contienen "bot". El valor de -0.64 entre "hora" y "recuento" indica una correlación negativa moderadamente fuerte, lo que sugiere que a medida que avanza el día, la cantidad de tweets que contienen "bot" tiende a disminuir. Mientras que los Tweet de usuarios también, sin embargo su baja correlación indica que las horas de publicación también tienden a bajar pero no dependen tanto de la hora.

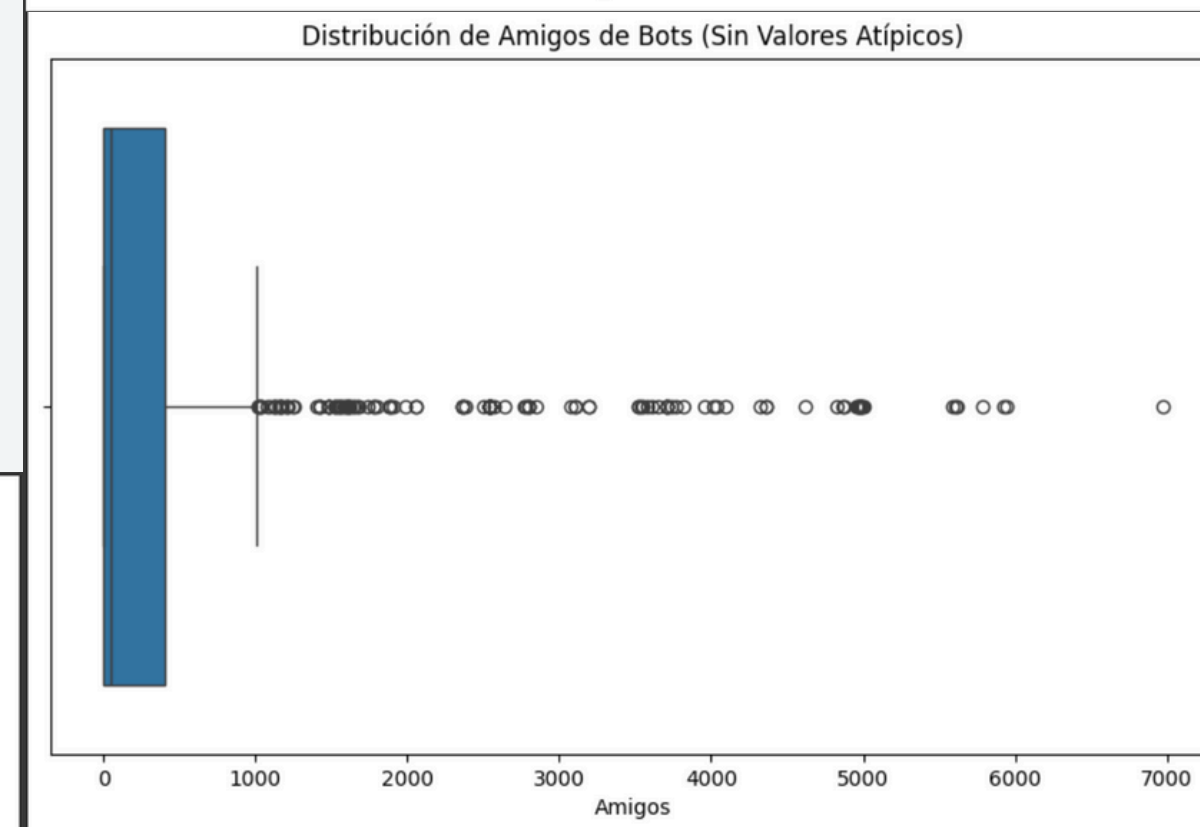
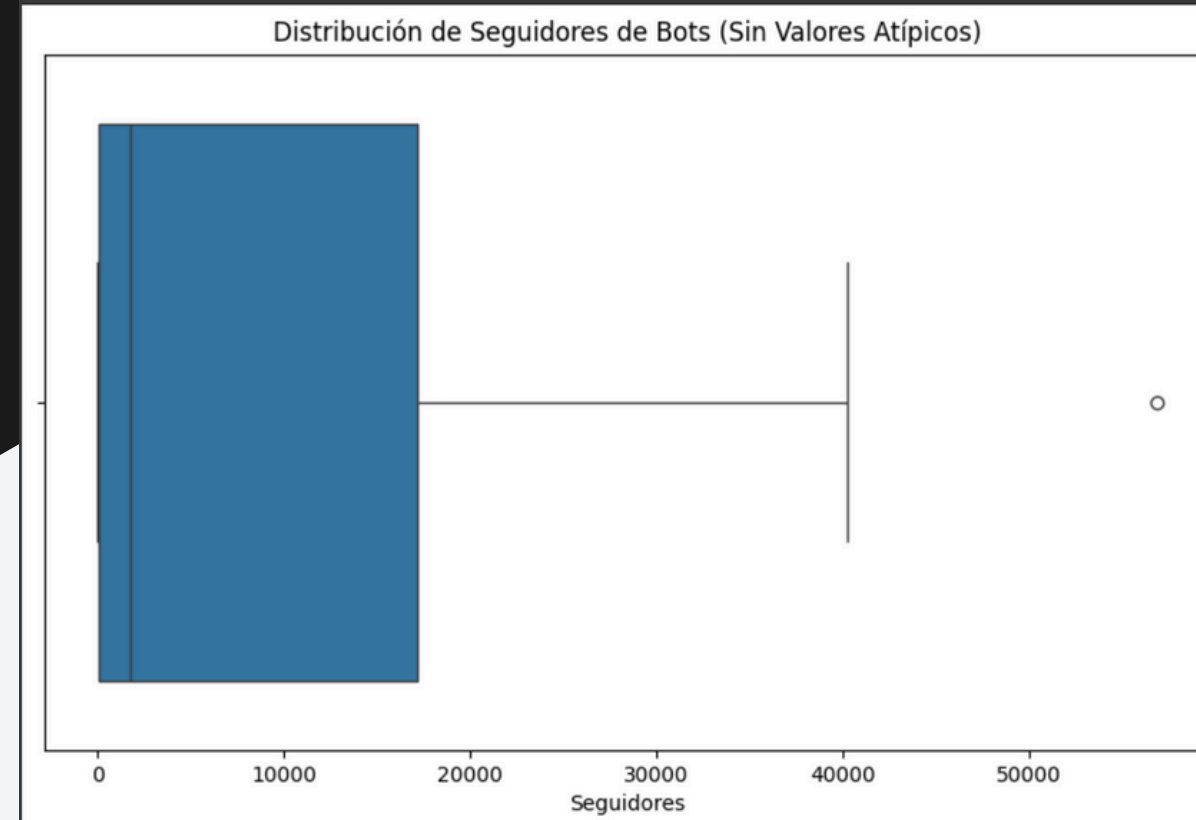
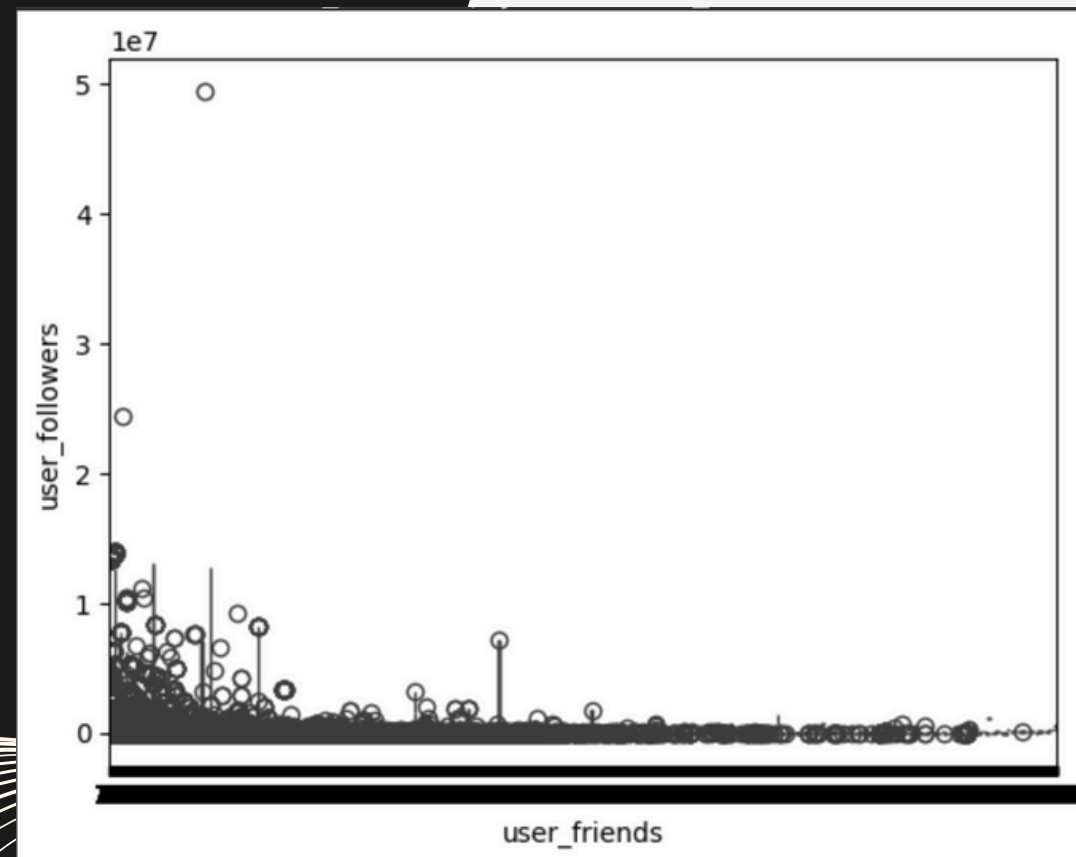


Este mapa de correlación analiza la relación entre la hora específica (minuto a minuto) y el recuento de tweets que contienen "bot". El valor de -0.1 entre "hora" y "recuento" muestra una correlación muy baja, indicando que no hay una relación lineal significativa entre la hora específica del día y el número de tweets mencionando "bot". Esto sugiere que la distribución de tweets relacionados a bots suelen tener horarios en punto, a diferencia de los usuarios normales que tienden a publicar en cualquier momento

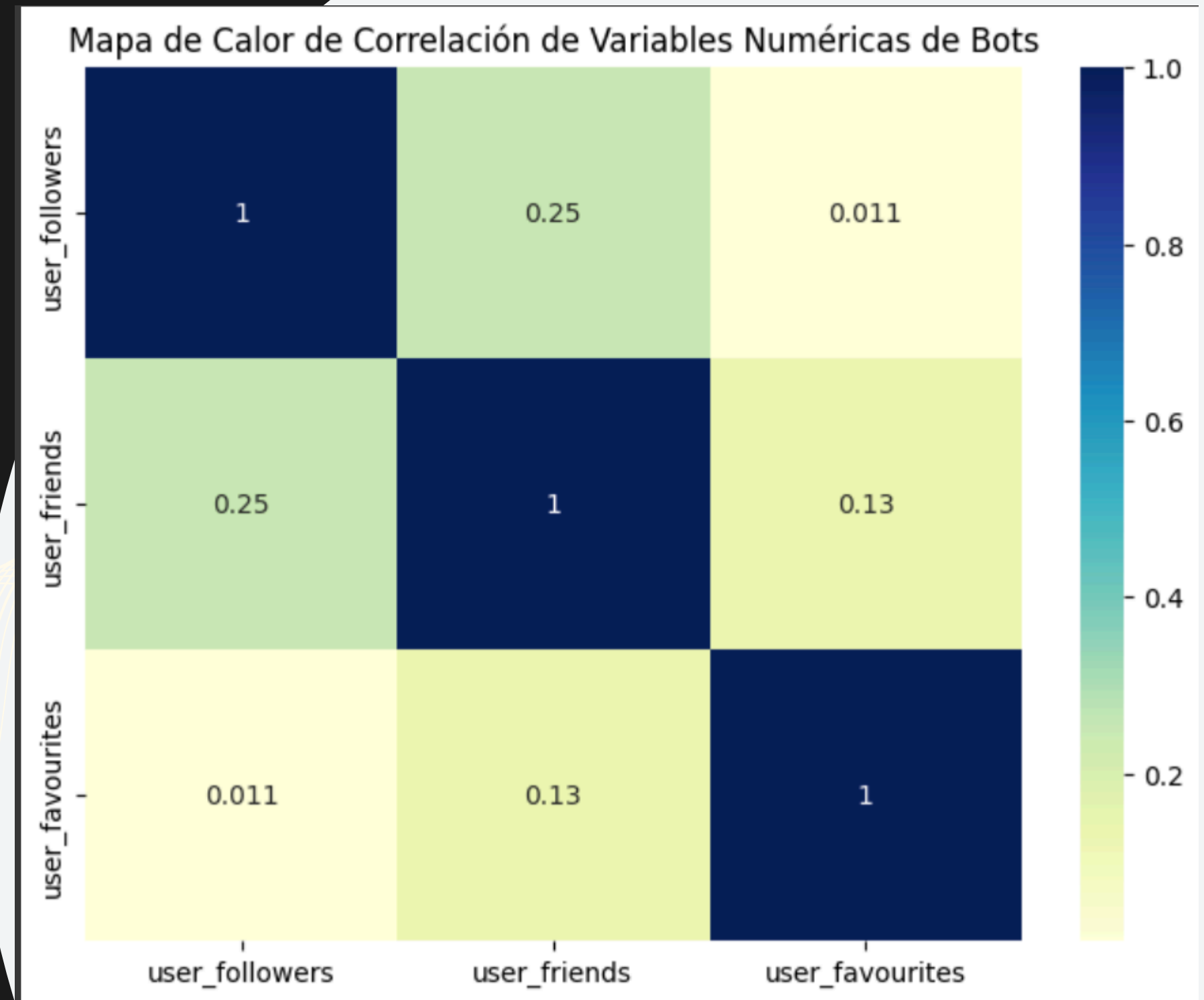


En estos diagramas se puede observar que en la base de datos, hay más bots con menos seguidores y tienen menos amigos. Se limpiaron los datos, quitando aquellos con menos de 100000 seguidores y 8000 amigos.

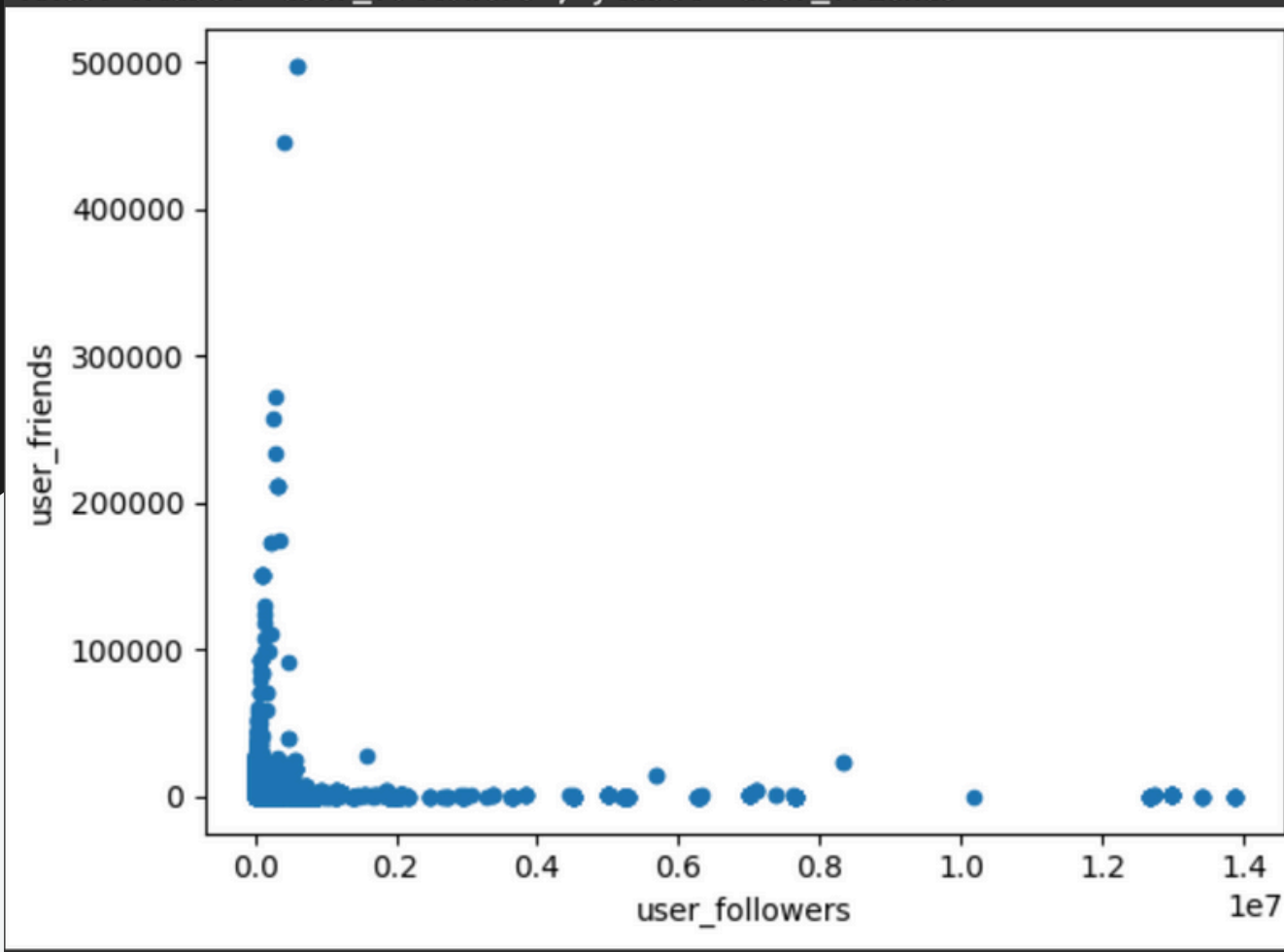
Asimismo, se puede observar que hay muchos valores atípicos en la gráfica de amigos, por lo que muy pocos bots tienen grandes cantidades de amigos. También, la mayoría de bots tienen muchos amigos pero pocos seguidores.



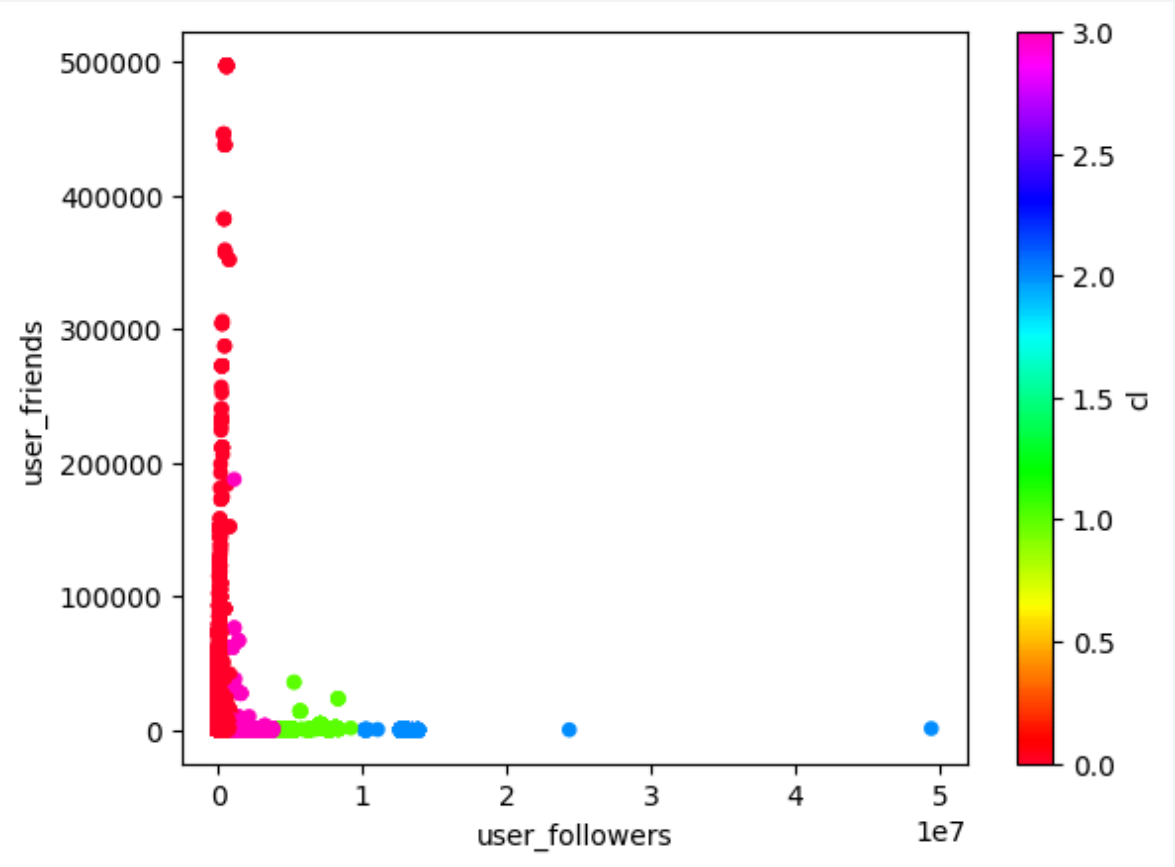
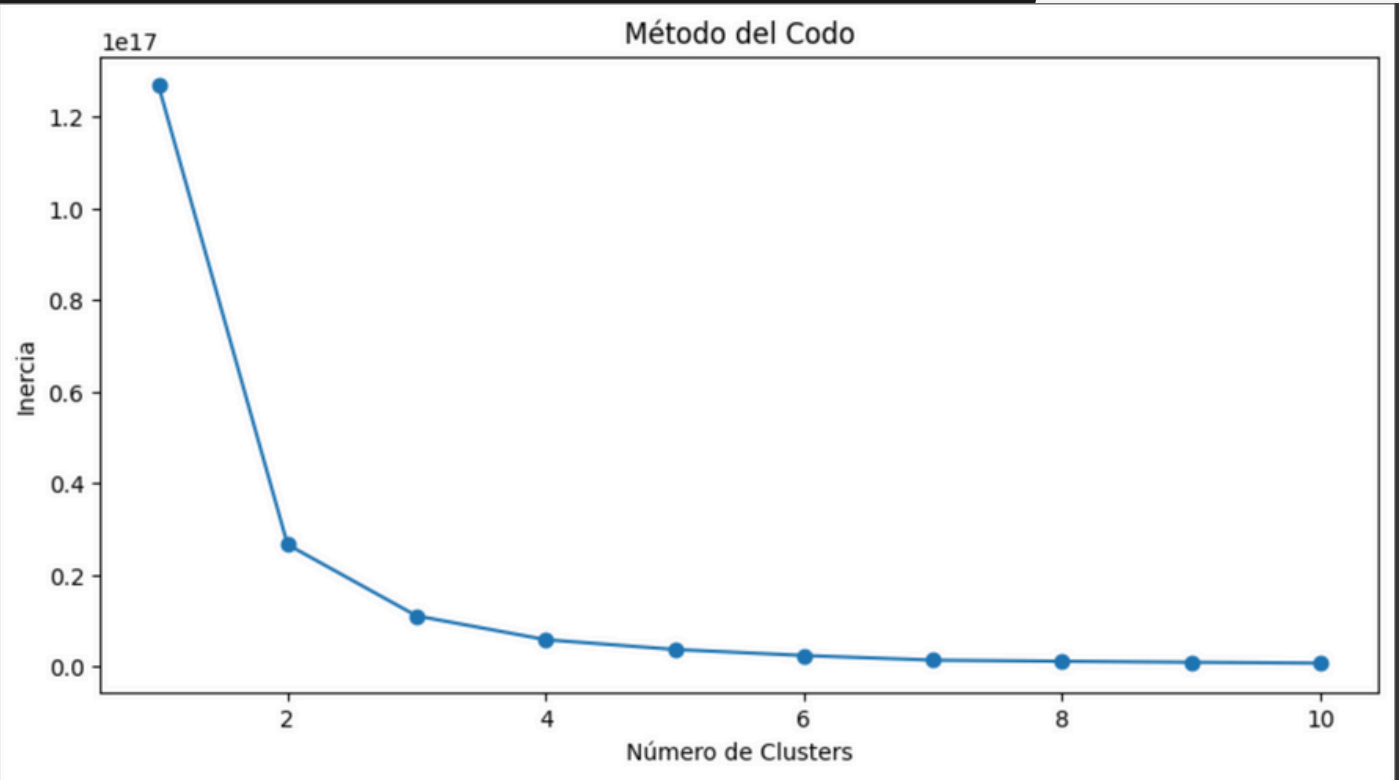
El mapa de calor muestra correlaciones débiles entre el número de seguidores, amigos y favoritos de usuarios en una red social, sugiriendo que estas variables son relativamente independientes entre sí. La correlación más notable, aunque todavía débil, es de 0.25 entre seguidores y amigos, indicando que usuarios con más seguidores tienden ligeramente a seguir a más personas.



Los gráficos presentan un análisis de la relación entre seguidores y amigos de usuarios en twitter, utilizando técnicas de clustering para identificar grupos con patrones de conexión similares. El método del codo sugiere que cuatro es el número óptimo de clústeres, reflejando una diversidad significativa en las redes de contactos entre los usuarios, desde aquellos con pocas conexiones hasta usuarios extremadamente conectados.



Centros de los clusters:
1.78572048e+04 2.14188283e+03
6.23214067e+06 7.21957189e+02
1.31393268e+07 2.40845960e+02
1.90965535e+06 1.53565129e+03



Cabe resaltar que todos los datos usados en este análisis fueron limpiados para que únicamente tenga en su username, hashtags y source la palabra "Bot".

CONCLUSIONES



Los usuarios de Twitter bots, casi no tienen relación entre sus amigos, seguidores y favoritos. Por lo que la gente se da cuenta que son bots y no tienen el impacto de una cuenta llevada por humanos.

Los



La actividad de los bots en Twitter muestra una fuerte correlación negativa con la hora del día, mientras que la actividad de los usuarios comunes tiene una correlación más débil. Esto indica que los bots son más propensos a publicar durante las horas pico, mientras que los usuarios comunes son más activos a lo largo del día.