

DataCompany



The Bridge Tech

Caso de uso

@TheBridge_Tech: que desea desarrollar una monitorización de las redes sociales para medir el impacto de su marca y acciones comerciales.

Tweets desde el 19 Septiembre hasta el 5 de diciembre

- a. Id del mensaje
- b. Cuerpo del texto del mensaje
- c. Fecha del tweet
- d. Id del autor
- e. Nombre del autor
- f. Nombre de usuario del autor
- g. Métricas públicas del tweet (retweet, reply, like, quote)

Datos



Csv con datos de tuits de febrero a mayo de 2023

Csv con datos correspondientes al usuario del primer dataset

Csv con datos de tuits de enero a mayo de 2022

	date	user	tweet	likes	rt	replies	user_id	nombre
0	2022-05-19	TheBridge_Tech	¡Qué poco queda para el próximo 🔥 Burn Monday 🔥!...	0	2	0	1162694149956603904.0	The Bridge
1	2022-05-18	TheBridge_Tech	Esta es la historia de Isis 😊 que decidió hacer...	0	0	0	1162694149956603904.0	The Bridge
2	2022-05-18	TheBridge_Tech	Gracias @victorianol y equipo por el análisis ...	0	0	0	1162694149956603904.0	The Bridge
3	2022-05-18	Bcasapp	¿Cómo combatir el paro juvenil 🤔? #tipdeldia En...	8	4	1		
4	2022-05-17	TheBridge_Tech	¿Con ganas de probar si un empleo tecnológico ...	2	2	0	1162694149956603904.0	The Bridge
...
272	2023-04-28	ProductoActual1	@TheBridge_Tech @EDEMempresarios @EDEM_uni ¿Po...	0	0	1	1613113784926707712.0	Manu
273	2023-05-02	jorgegrev	🇪🇸 HOY #OpenDay de los BBK Bootcamps by The Bri...	0	0	0	2402179123.0	Jorge González Revilla
274	2023-05-05	jorgegrev	🎂🎉 ¡Este mes se cumplen 10 años del primer #Be...	2	1	0	2402179123.0	Jorge González Revilla
275	2023-05-11	Madrid4Refugees	Join our #RefugeesCode Demo Day & networki...	5	5	1	3588258623.0	Madrid For Refugees
276	2023-05-19	a3sec	Ayer visitamos a los alumnos de @TheBridge_Tec...	3	1	0	987047282.0	A3Sec

277 rows × 8 columns

SQL



```
# Ruta donde se creará la base de datos 'tweets_usuarios'
database_path = r'C:\Users\bego\Desktop\DS\bootcampPrueba\alumna\DS_TheBridgeBBK_SBIL2023\3-Machine_Learning\4-NLP\Entregas\sql\tweets_

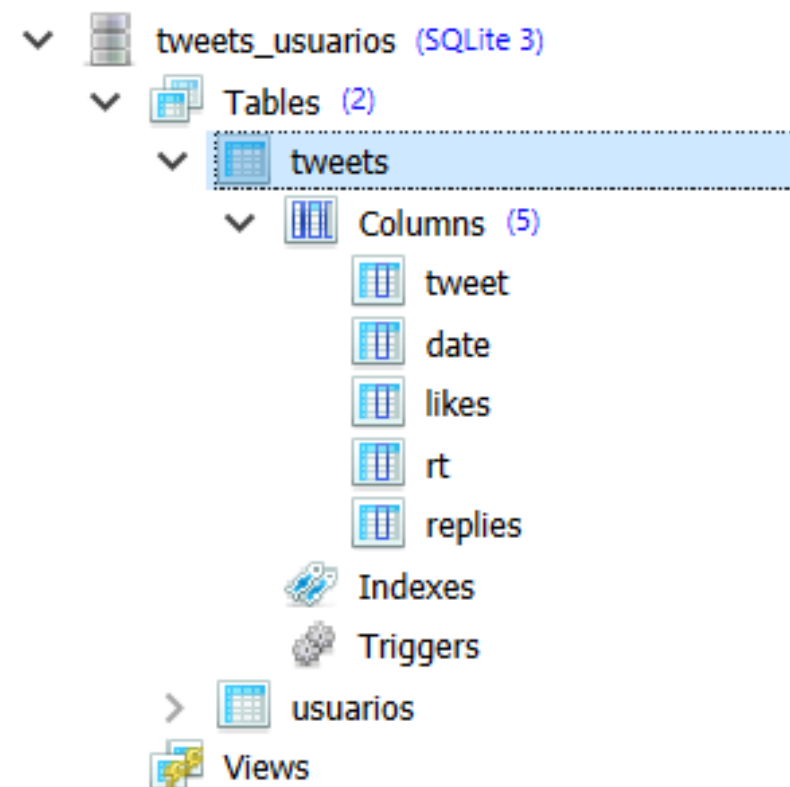
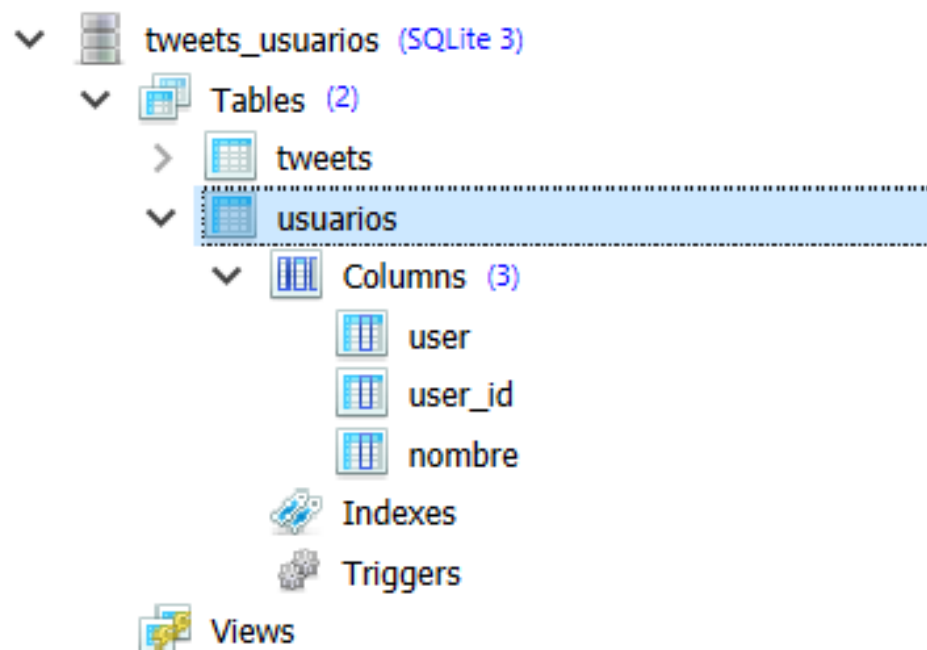
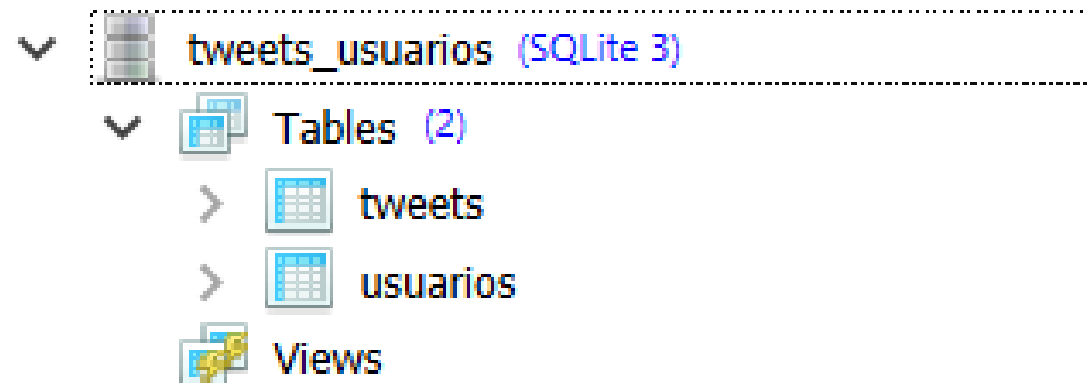
# Conectar a la base de datos (se crea el arch)
conexion = sqlite3.connect(database_path)

# Guardar los datos en la base de datos
with conexion:
    # Guardar usuarios en la tabla 'usuarios'
    t_concat[['user', 'user_id', 'nombre']].drop_duplicates().to_sql('usuarios', conexion, if_exists='append', index=False)

    # Guardar tweets en la tabla 'tweets'
    t_concat[['tweet', 'date', 'likes', 'rt', 'replies']].to_sql('tweets', conexion, if_exists='append', index=False)

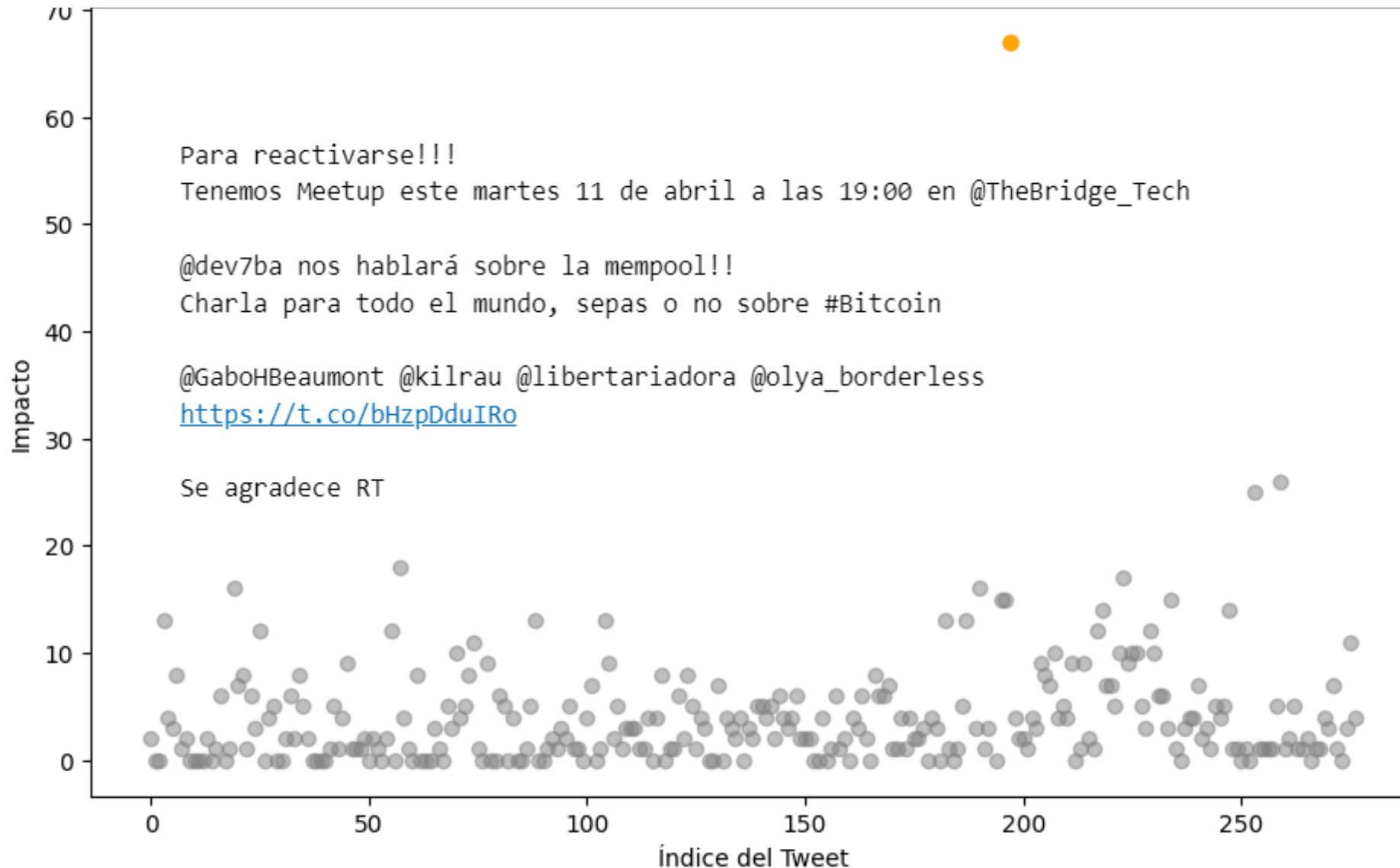
# Cerrar la conexión
conexion.close()
```

SQL



Tweet con mayor repercusión social

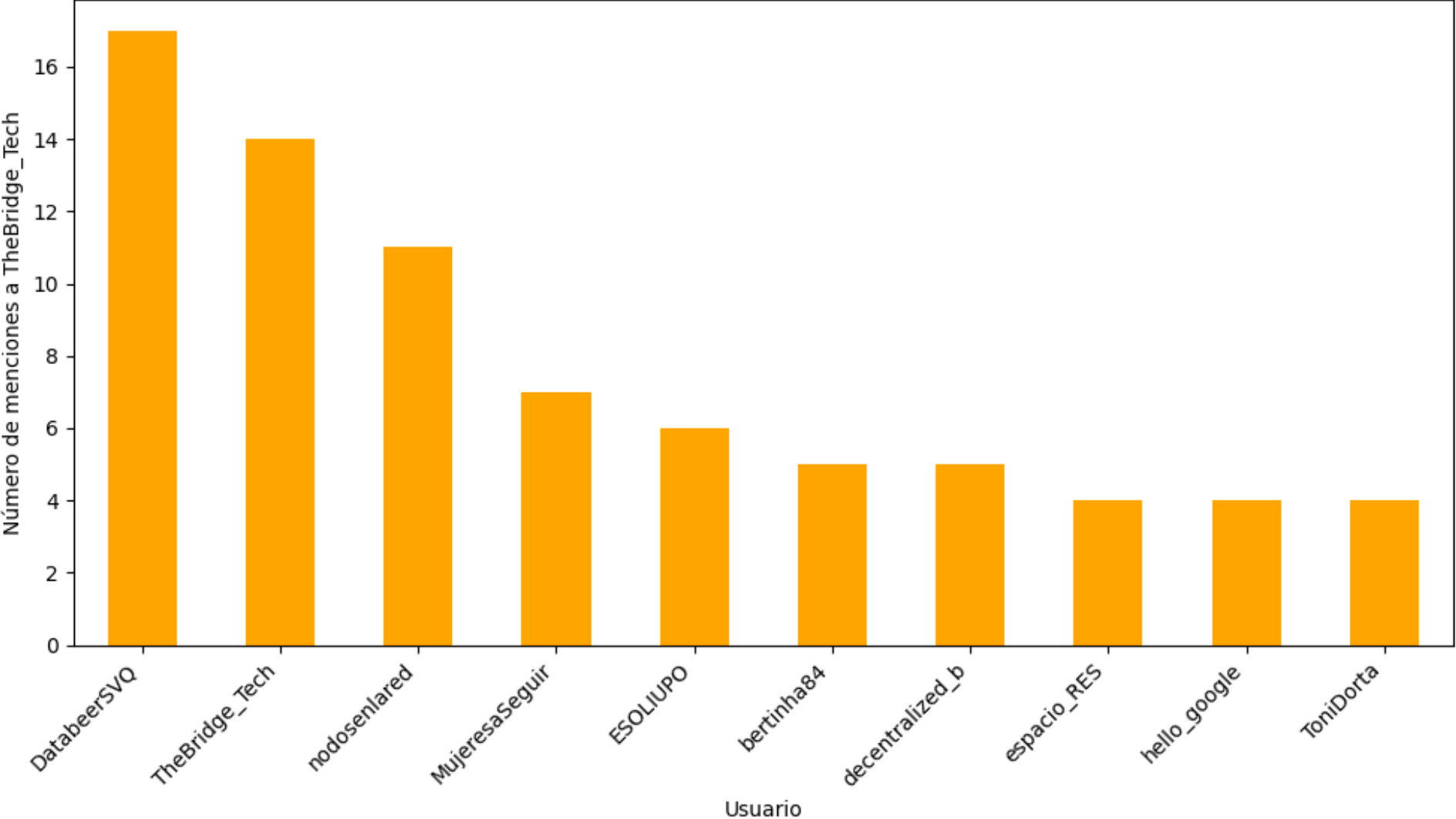
Análisis



Usuario que más menciona a la escuela

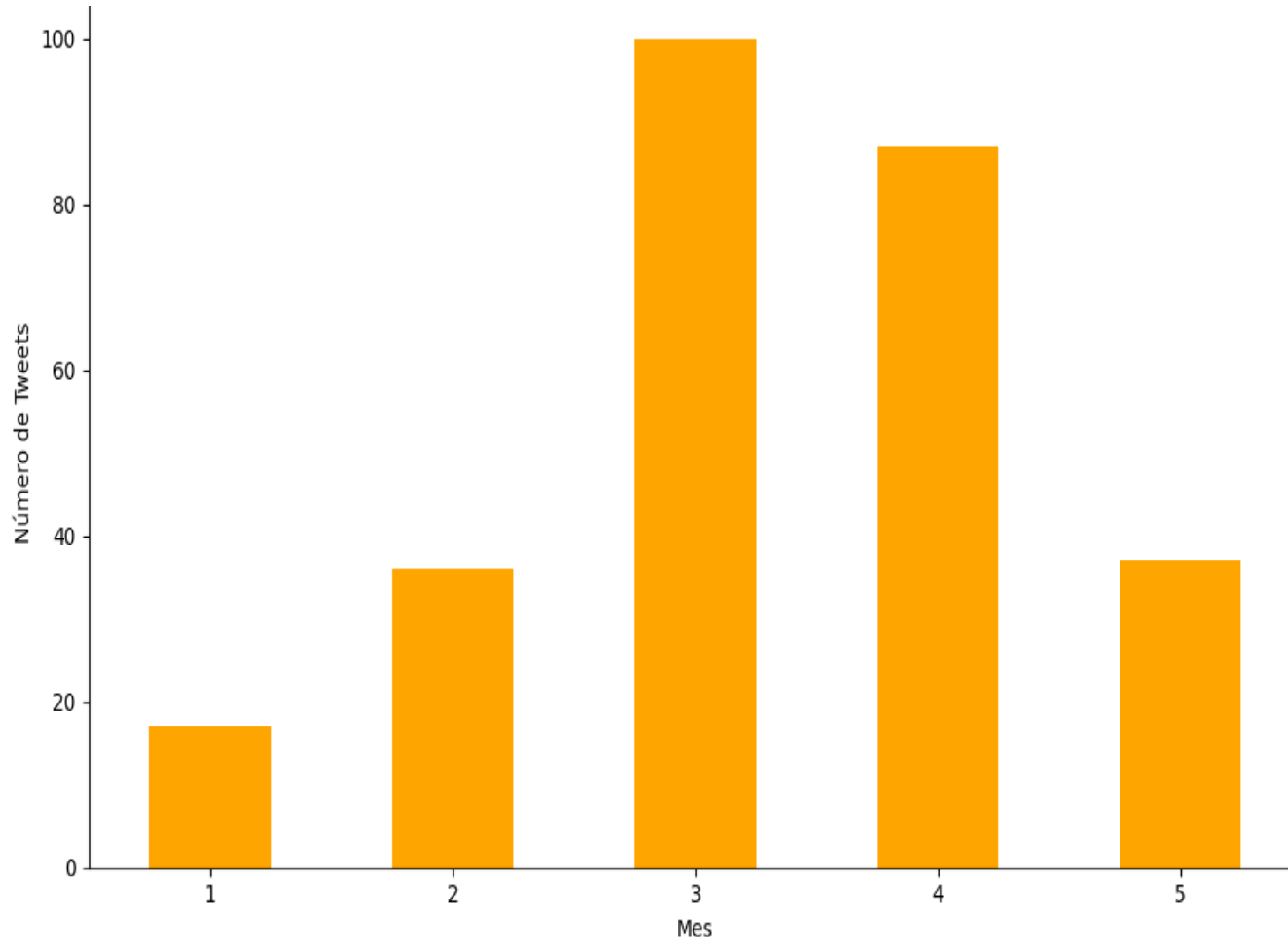
Análisis

Top 10 Usuarios que mencionan a TheBridge_Tech



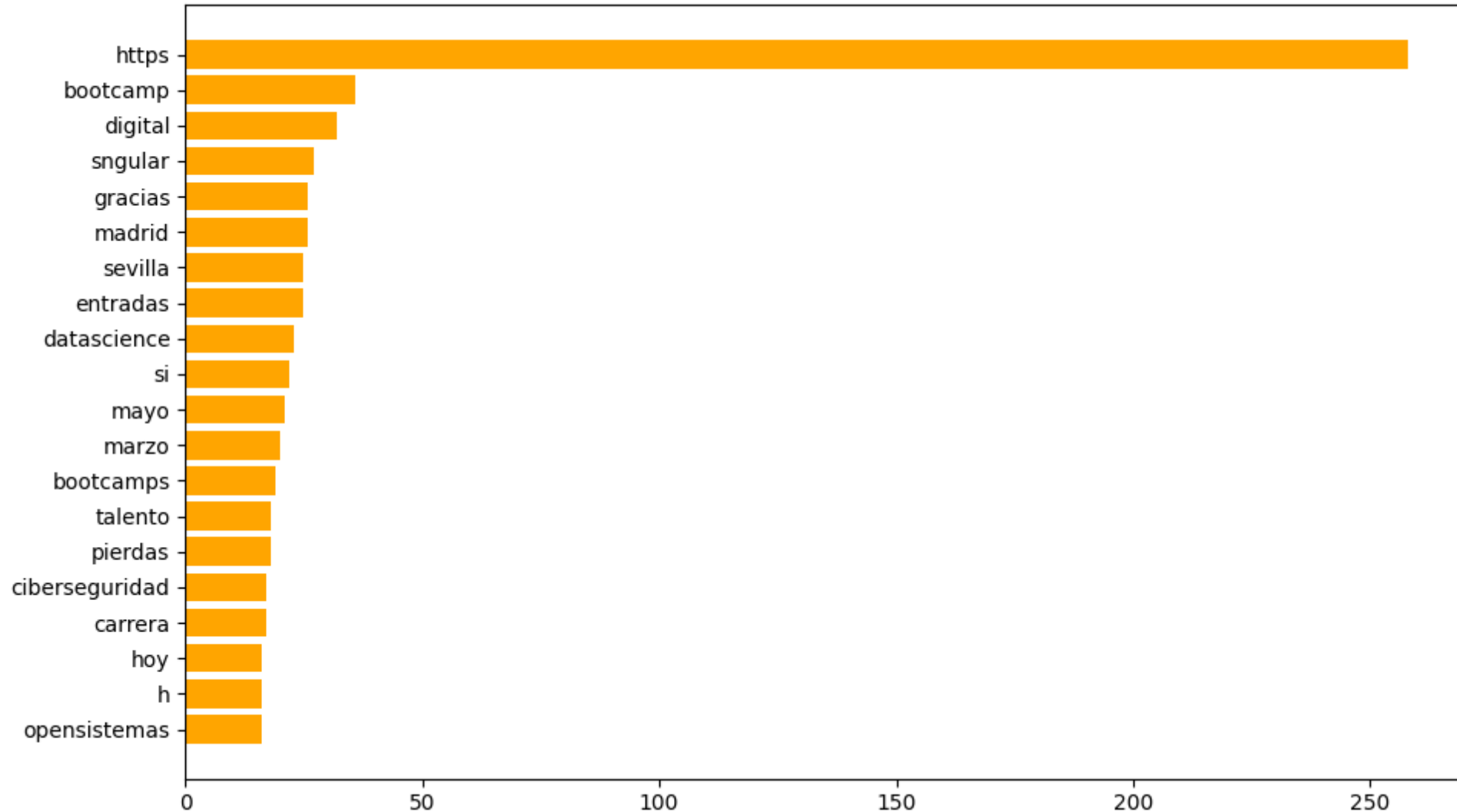
Mes con el mayor número de tweets

Análisis

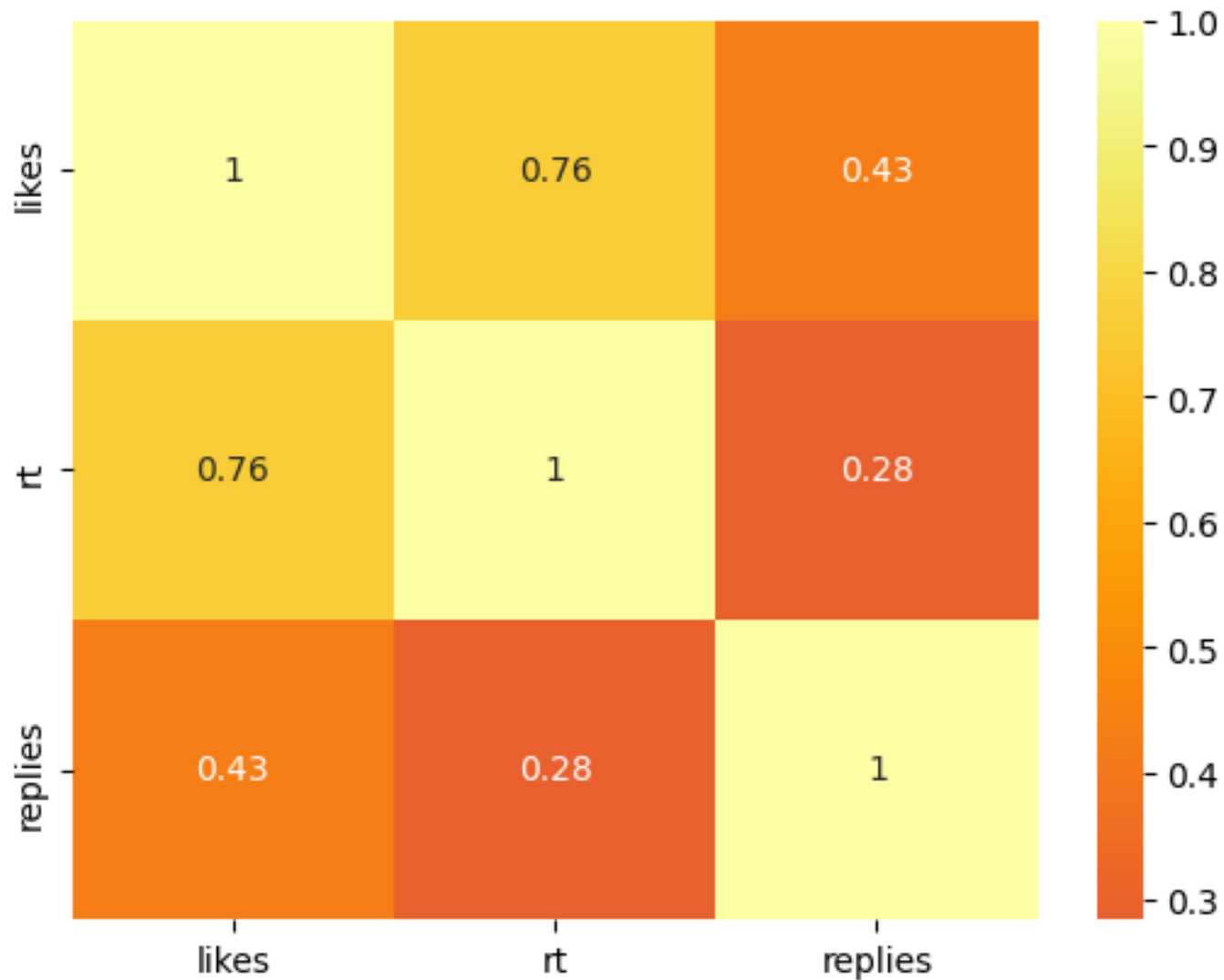


Palabras son más frecuentes

Análisis



Correlación matemática entre las métricas



Análisis

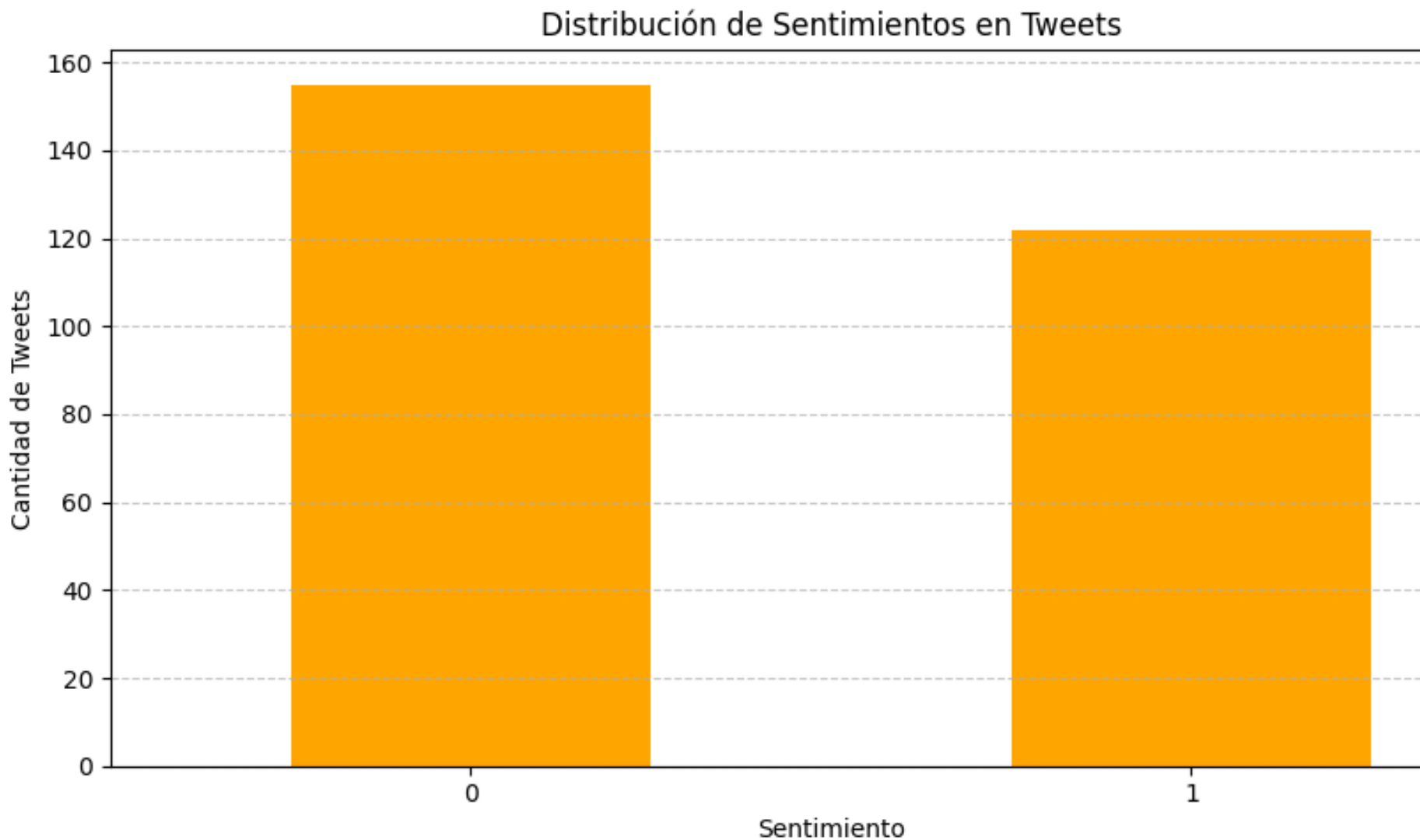


Análisis



- El usuario 'DatabeerSVQ' destaca por su alta frecuencia de menciones a 'TheBridge_Tech', podría indicar algún tipo de conexión entre este usuario y la escuela.
- Marzo fue el mes con más actividad, seguido de abril, lo que podría ser indicativo de inicio de cursos, eventos relevantes, lanzamientos o campañas durante ese periodo.
- La presencia recurrente del término "https" indica la inclusión de enlaces en los tweets. No hay palabras que en sí mismas sean negativas en el top 20.
- Se identifica una correlación significativa entre 'likes' y 'rt' (retweets), indicando que los tweets populares tienden a recibir más retweets. Sin embargo las otras correlaciones entre 'likes', 'rt' y 'replies' son más débiles, lo que indica relaciones menos directas entre estas métricas.

Modelo



sentimiento_predicho

0	155
1	122

Modelo



- Feature engineering, inclusión de nuevas características relevantes, buscar la información que faltaba (NaN).
- Limpiar y procesar adecuadamente los textos eliminando stop words, normalización de texto, corrección de errores tipográficos...
- Probar con otros modelos, Random Forest, SVM, Redes Neuronales... y ajustar hiperparámetros.
- Aumentar el conjunto de datos con más ejemplos.
- Validación cruzada y métricas de evaluación adecuadas para comprender mejor cómo se desempeña el modelo.
- Técnicas de ensemble como Bagging, Boosting o modelos apilados para combinar múltiples modelos y mejorar la generalización.
- Estudiar los errores cometidos por el modelo.



