

Hey, Banco 2024

Sentiment Analysis

Adrián Mateos
Sebastián Miramontes
Diego Gutiérrez
Miguel González
DSMM team

Problem identification	The team clearly and thoroughly identifies the problem to be solved, demonstrating a deep understanding of its root causes and potential impact (15 points)
Data exploration	The team shows a thorough exploration of the data, including resolving doubts and using additional sources of information where appropriate (15 points)
Idea generation	The team generates 3-4 high-quality ideas as a team, each of which demonstrates creativity and potential value to the problem at hand (20 points)
Solution development	The team chooses the idea that is considered to generate the most value, develops it into a viable solution, and considers potential MVPs (15 points)
Technical quality of model	The team demonstrates technical excellence in the development and execution of the model, showing a thorough understanding of relevant concepts (5 points)
Business case preparation	The team prepares a comprehensive and well-reasoned business case, including clear recommendations for how the solution can be implemented (10 points)
Viability of replicability	The team develops the solution with an eye towards its long-term viability, considering scalability, maintainability, and potential obstacles (10 points)
Insights	The team provides actionable insights into the problem at hand, demonstrating a thorough understanding of its nuances and potential solutions (10 points)

Overview

1. Identificación del problema
2. Exploración de datos
3. Generación de ideas
4. Desarrollo de solución
5. Calidad técnica del modelo
6. Caso aplicado a negocios
7. Viabilidad de replicación
8. Insights



Identificación del problema

Problematica: Oportunidad para clasificar interacciones en Twitter; relacionar sentimientos presentes y analizar el mensaje transmitido.



Pasos para solucionarla:

- 01 Se brindó un DataSet con tweets que mostraran interacciones
- 02 Llevar a cabo análisis de sentimientos para ver comportamiento
- 03 Creación de gráficos y análisis Exploratorio de los resultados
- 04 Construir insights y recomendaciones

Exploración de datos

- Limpieza de datos: estandarización, descartando acentos y minúsculas.
- Valores nulos en tiempo irrelevantes.
- Exploración manual de datos
- Aceptar emojis, dado que describen una emoción.
- Primera búsqueda: implementación de algoritmos en base a listas de emojis, palabras o signos claves para cada sentimiento.
- Segunda búsqueda: implementación de librerías con análisis de sentimientos para palabras en español (PYSentimiento y Sentiment Analysis Spanish).

	date	time	tweet
0	2023-01-01	NaN	resuelto, muchisimas gracias , excelente servi...
1	2023-01-02	NaN	muchas gracias, espero su dm
2	2023-01-02	NaN	muchas gracias!
6	2023-01-03	NaN	gracias 😊
7	2023-01-03	NaN	muchas gracias 😊
...
803	2024-04-17	04:45:42	gracias 😊
804	2024-04-18	17:28:23	gracias
805	2024-04-18	15:59:11	los amo! gracias ya pude entrar a mi app salud...
807	2024-04-21	15:21:43	ahora confirmo, gracias estimado
808	2024-04-23	22:52:05	😊 de acuerdo!! muchas gracias.

Generación de ideas



Clasificación de modelos mediante fuentes externas

Clasificación de modelos mediante algoritmos

Clasificación mediante fuentes modelos de clasificación

Análisis de sentimiento por línea de tiempo

Análisis de sentimiento por palabras clave

Desarrollo de solución

NLPs y Manual: Algoritmos basados en 2 distintas librerías NLPs para el reconocimiento del sentimiento y emoción, incluyendo emojis y palabras clave para la clasificación de sentimientos.

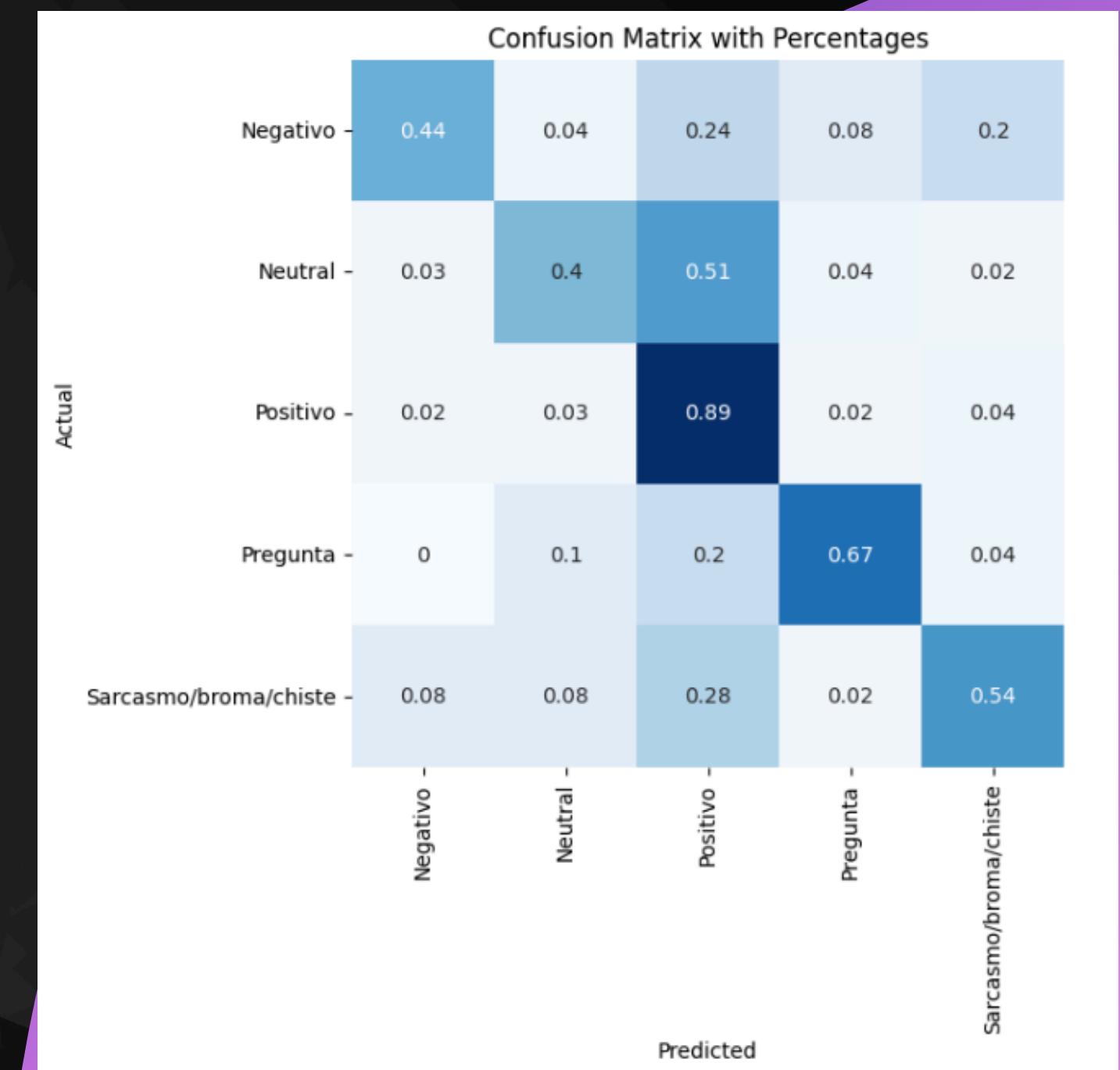
SVC: Clasificador usando las librerías NLPs para la predicción de sentimientos.

Calidad Técnica del Modelo: NLPs y Manual

Valores iguales: 603, Porcentaje: 74.26108374384236%

Valores diferentes: 209, Porcentaje: 25.738916256157633%

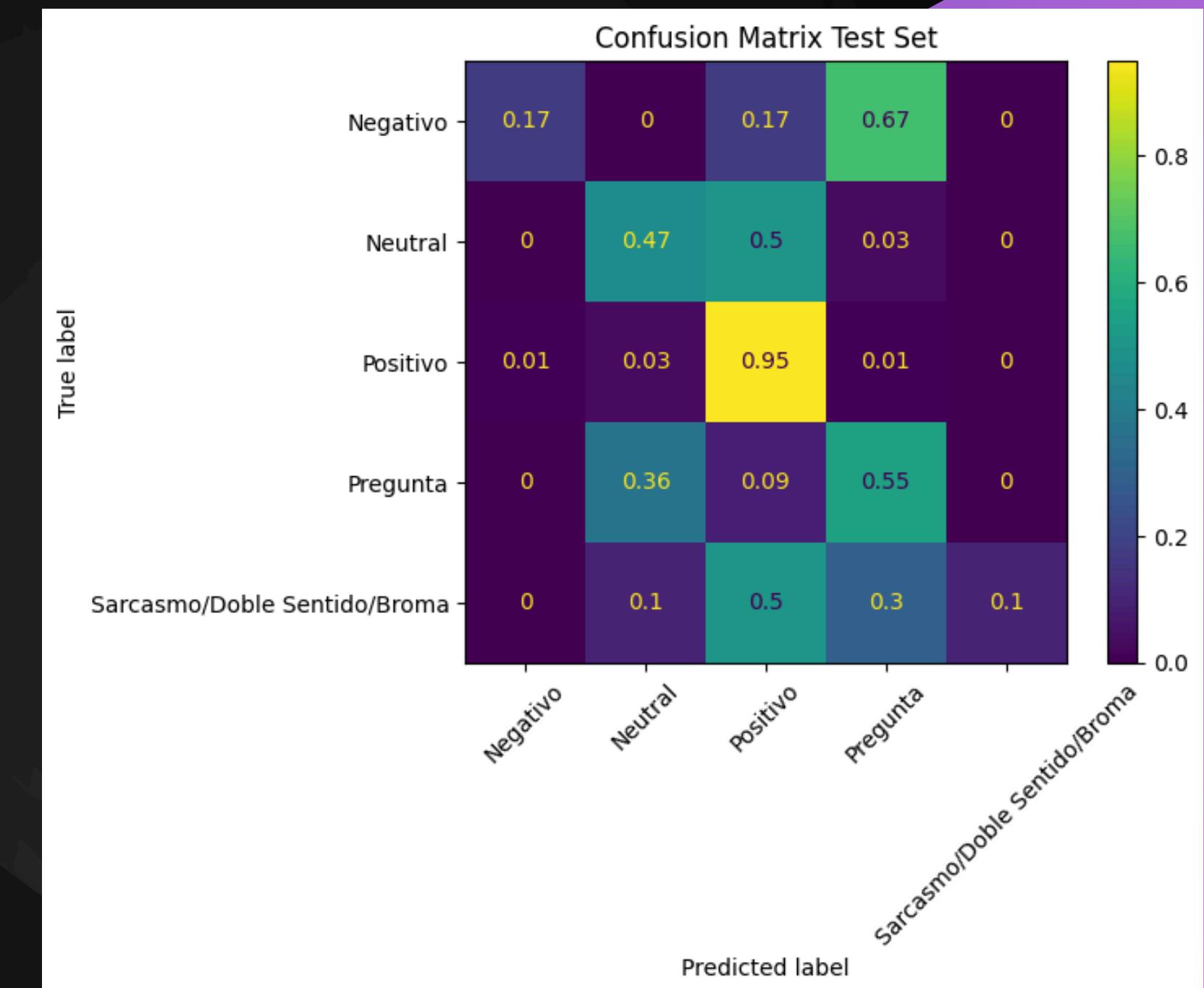
Este Modelo consistió en crear combinaciones entre las predicciones creadas a través de las librerías y las creadas por los algoritmos, hasta llegar a la solución con más valores reales posible.



Calidad Técnica del Modelo: SVC

Accuracy Modelo SVC(linear,C=1): 0.7484662576687117

Se utilizaron las columnas creadas con las NLPs en pasos anteriores y las combinaciones con las palabras y emojis establecidas, para realizar un modelo de “Support Vector Classifier” usando distintos hiperparámetros, creando un ajuste con grid search.



Business Case

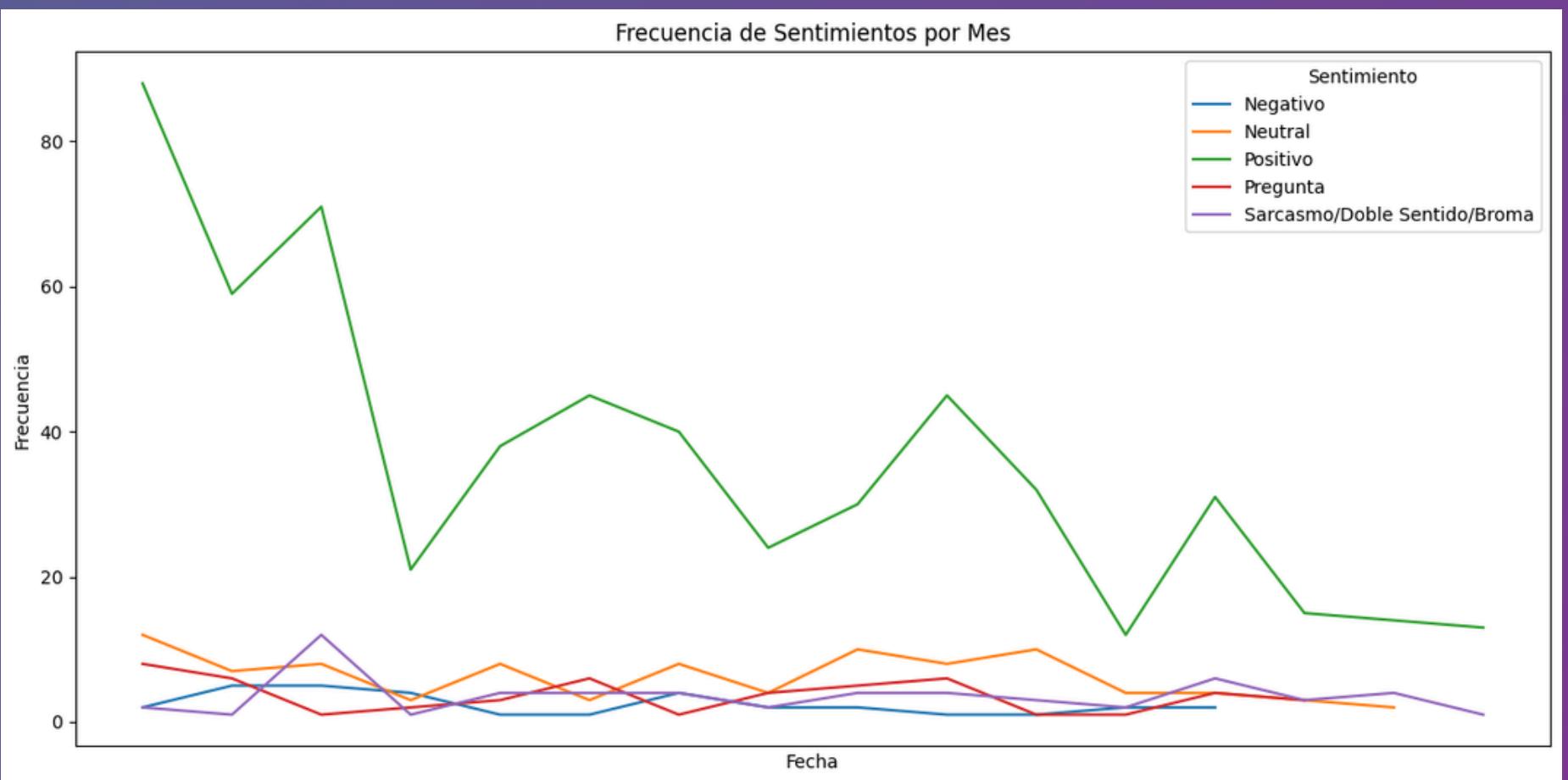
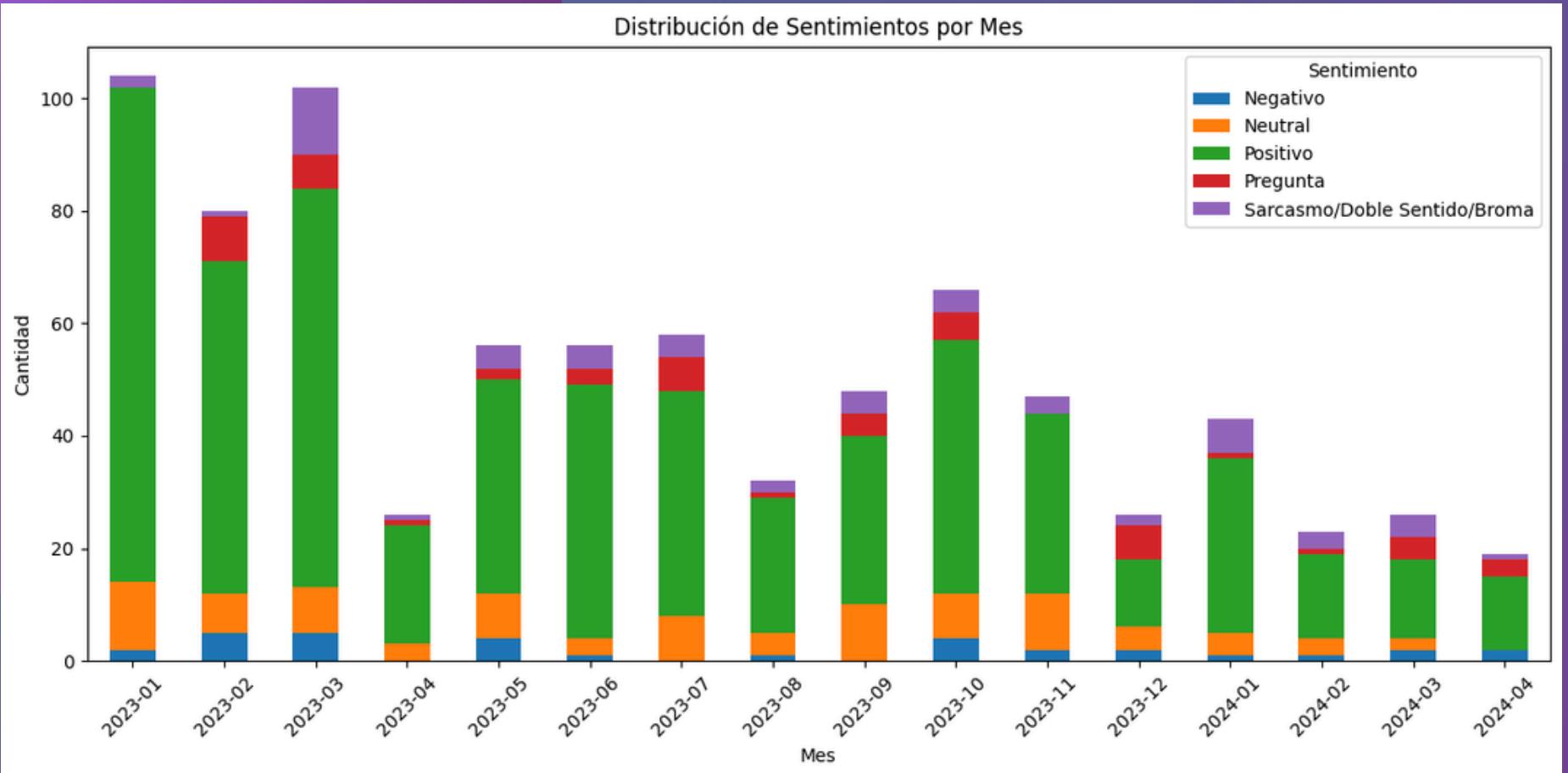
- Probar la efectividad de campañas de marketing
- Medida de desempeño para personal de servicio al cliente
- Analizar tendencias de mercado
- Percepción pública
- Percepción pública del servicio al cliente

Seguimiento a futuro

- **A medida que aumenta el número de datos:**
- **Mayor precisión:** Mejores modelos para aprender más acerca las relaciones entre las palabras y emociones, lo que permitirá un porcentaje de identificación más alto; insights más profundos.
- **Mantenibilidad:**
- **Documentación:** La documentación completa y apropiada del código y los procesos realizados; facilita la comprensión por parte de otros desarrolladores.
- **Actualización de datos:** Asegurar que el modelo se alimente con datos nuevos y relevantes de manera regular.
- **Obstáculos Potenciales**
- **Calidad de los datos:** La precisión del modelo depende de la calidad de los datos de entrada.
- **Evolución del lenguaje:** El lenguaje evoluciona constantemente, por lo que el modelo debe ser actualizado para adaptarse.

Insights

- DataSet carece de estructura para entrenar adecuadamente un modelo predictorio.
- Existía un desbalanceo en los datos; la mayoría de interacciones fueron clasificadas como ‘positivas’
- Buena señal: mayoría de interacciones con comunidad de Twitter son de respuestas positivas y agradecimientos
- Capturar una muestra con mayores registros de interacciones.
- Generar dinámica en las interacciones con la comunidad para ver su comportamiento; mayoría de resultados es atención al cliente.
- Identificar los tópicos principales de los que hablan las interacciones ‘negativas’ para buscar patrones.

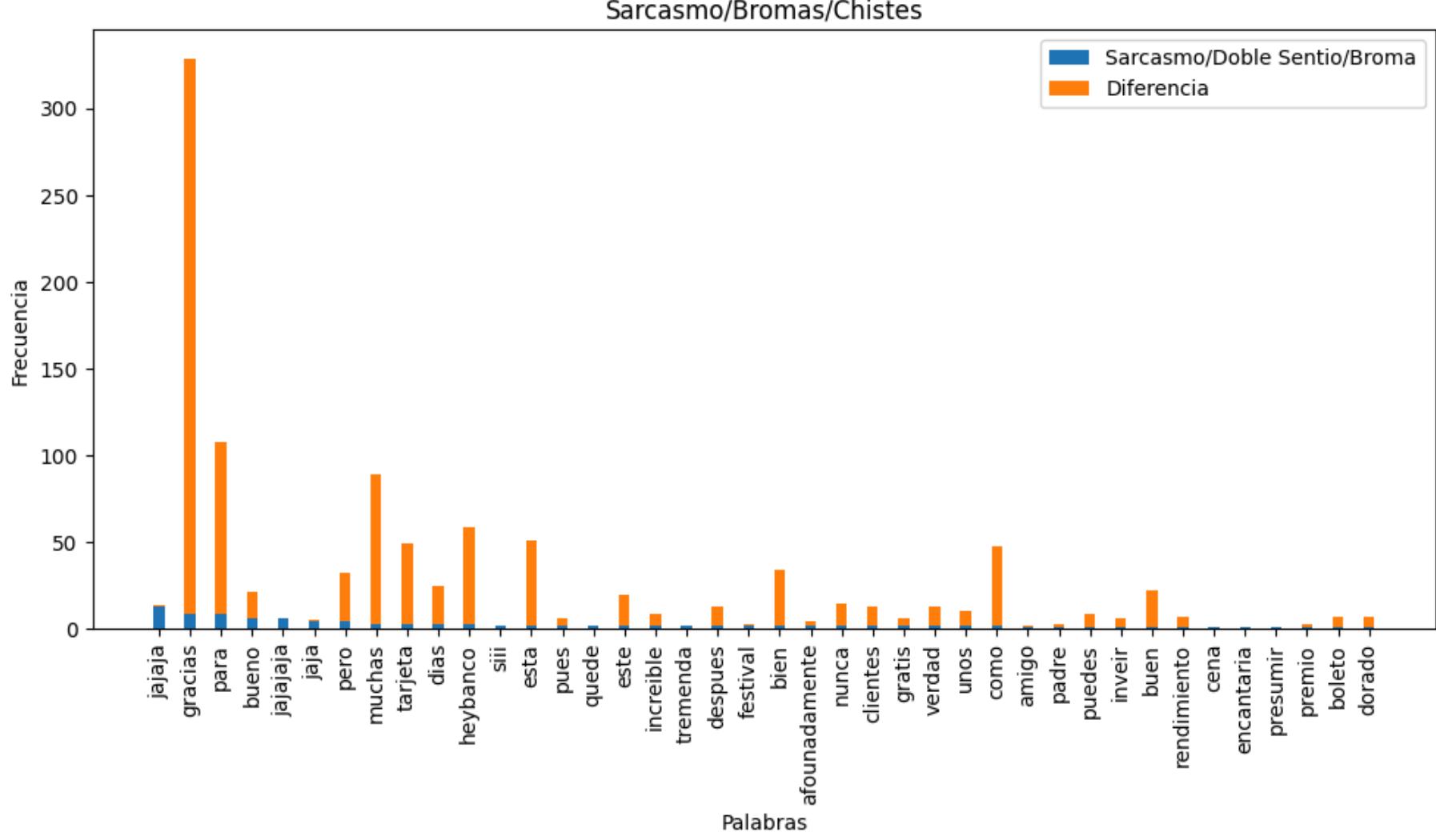


Decaimiento en el engagement

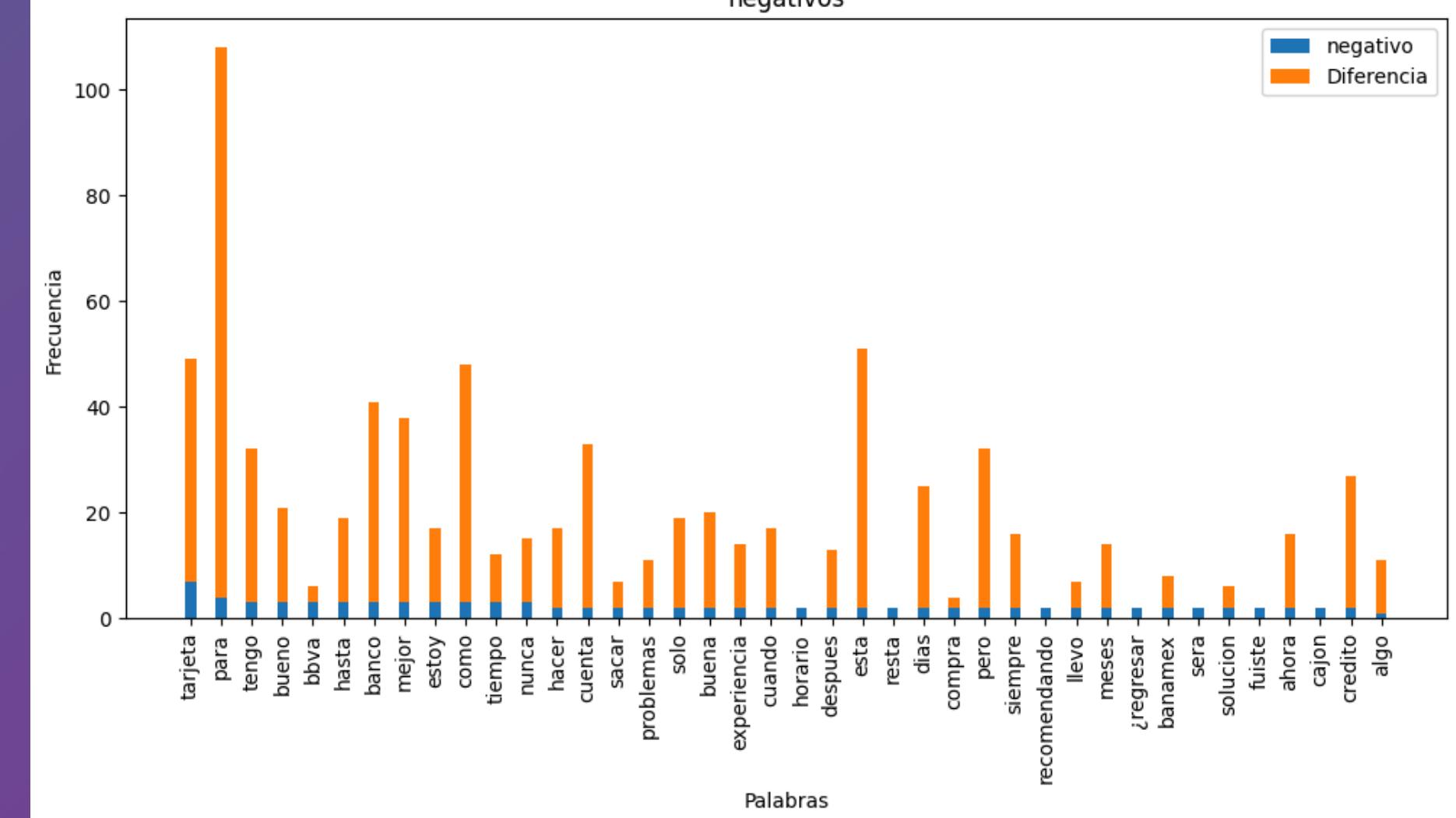
A pesar de que en todos los meses se puede observar que la gran mayoría de las interacciones son positivas, también se puede observar que al pasar del tiempo la cantidad de las mismas va disminuyendo.

Se debe abordar este problema, ya que las interacciones pueden ayudar a mantener una buena imagen del banco, además de generar una mayor cantidad de vistas y aumentar el conocimiento sobre él

Sarcasmo/Bromas/Chistes



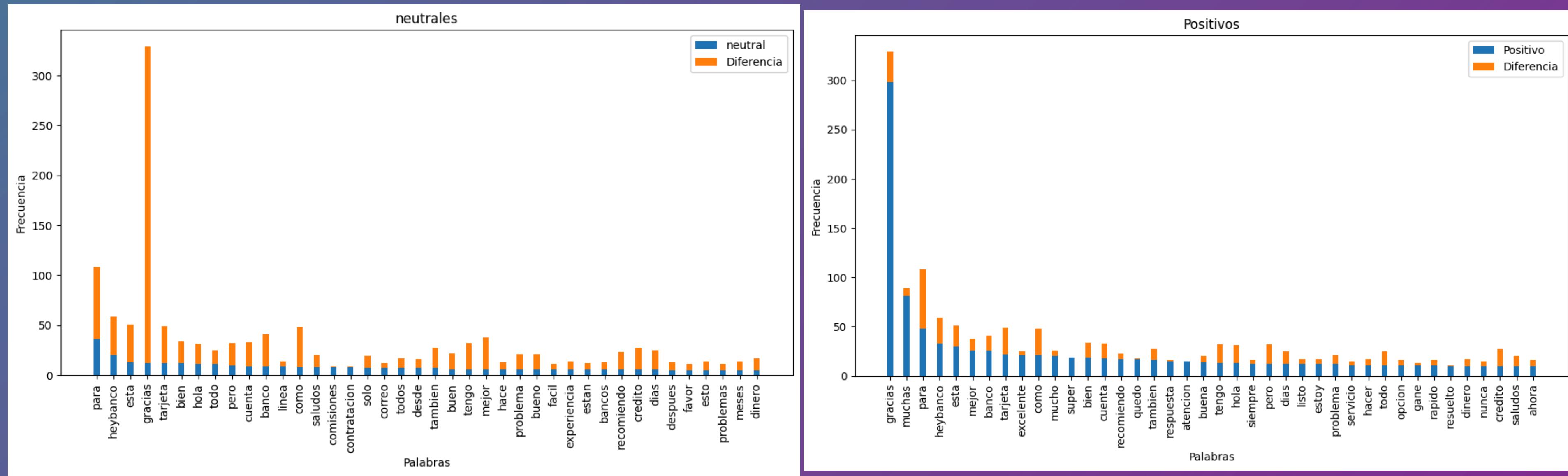
negativos



Palabras destacadas (sarcasmo o negativos):

De acuerdo a los datos no se encontraron una gran variedad de datos sobresalientes en ninguno de los dos casos.

La más interesante fue la palabra “horarios” en los negativos, por lo que se podría considerar que las únicas veces en donde se comentó acerca de esto, fueron quejas.



Palabras destacadas (neutrales):

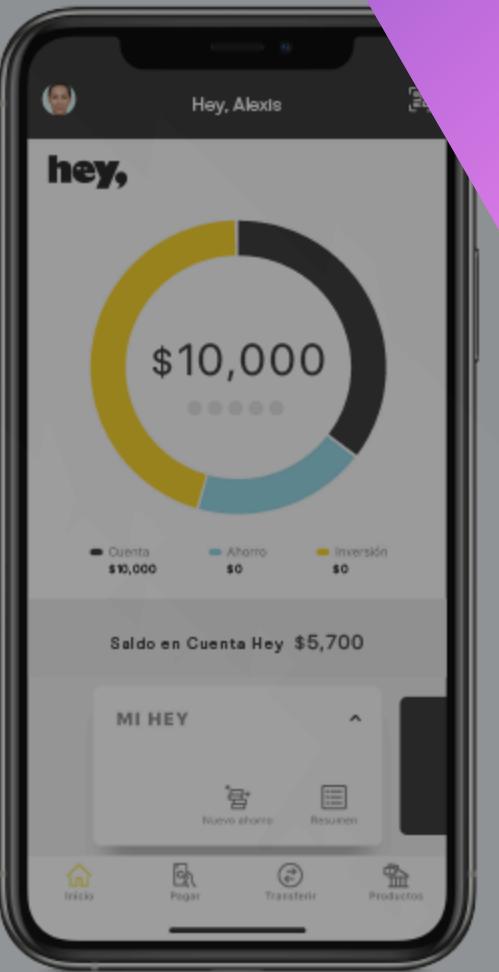
En las palabras neutrales destacadas encontradas se pueden observar: “comisiones”, “concentracion” y “correo”, por lo que no son aspectos negativos, pero podrían mejorar.

Palabras destacadas (positivos):

Las palabras mas destacadas se puede apreciar que son: “muchas” y “gracias”. Lo que sugiere que la mayoria de tweets positivos provienen de personas cuyas dudas han sido resueltas.

¡GRACIAS!

hey,
banco



DUDAS?