Actividad Evaluable: Mapas de calor y boxplots

**¿Hay alguna variable que no aporta información?**

En este caso, algunas de las columnas adicionales como Column1, Column2, Column3, y Column4 parecen no tener datos útiles (están llenas de valores NaN). Estas variables pueden eliminarse ya que no aportan información relevante para el análisis.

**Si tuvieras que eliminar variables, ¿cuáles quitarías y por qué?**

Eliminaría las siguientes variables:

1. **Column1, Column2, Column3, Column4**: Estas columnas contienen solo valores nulos (NaN), por lo que no aportan ninguna información útil.
2. **status\_id**: Aunque esta columna contiene identificadores únicos para cada post, no es relevante para el análisis de relaciones y métricas de interacción como comentarios, reacciones o compartidos.

**¿Existen variables que tengan datos extraños?**

No parece haber datos extraños en las variables numéricas clave como num\_reactions, num\_comments, num\_shares, num\_likes, etc., ya que todos los valores parecen razonables dentro de su contexto. Sin embargo, algunos valores pueden parecer atípicos, como aquellos con un número extremadamente alto o bajo de reacciones o comentarios. Estos podrían ser considerados "outliers", pero eso dependería del contexto del dataset y de si los números extremadamente altos representan algo excepcional o un error.

**Si comparas las variables, ¿todas están en rangos similares? ¿Crees que esto afecte?**

No todas las variables están en rangos similares, como se observa en los histogramas. Algunas variables como num\_reactions o num\_likes tienen rangos mucho mayores que otras, como num\_angrys, num\_sads, o num\_hahas, que tienden a tener valores mucho más bajos. Esto podría influir en el análisis, ya que las variables con mayores valores podrían dominar la interpretación de correlaciones o distribuciones.

Para evitar que una variable con un rango mucho mayor afecte el análisis, podríamos normalizar o estandarizar los datos antes de realizar más análisis.

**¿Puedes encontrar grupos qué se parezcan? ¿Qué grupos son estos?**

Se pueden observar grupos de variables que están correlacionados, como num\_reactions, num\_likes, y num\_shares. Las publicaciones con más reacciones también suelen tener más "me gusta" y más compartidos. Estos son indicadores de que algunas publicaciones generan más interacción que otras.

A través de un análisis de clustering (por ejemplo, K-means), sería posible identificar grupos o tipos de publicaciones que tienen un patrón similar en cuanto a interacciones, como los videos frente a las fotos, o publicaciones con una gran cantidad de reacciones en comparación con otras que tienen pocas.

**Conclusión de las gráficas y la correlación resultante:**

Las gráficas de distribución muestran una gran variabilidad en las interacciones de las publicaciones. Mientras que algunas publicaciones tienen muy pocas reacciones, otras tienen interacciones masivas, lo cual es característico de las plataformas de redes sociales.

La **correlación** entre num\_reactions, num\_likes, y num\_shares es alta, lo que sugiere que los posts con más reacciones tienden a ser más compartidos y reciben más "me gusta". Esta es una relación esperada, ya que los posts con más interacción tienden a generar mayor visibilidad y, por ende, más interacciones adicionales.

Sin embargo, algunas variables como las reacciones negativas (e.g., num\_angrys, num\_sads) tienen correlaciones débiles con las reacciones positivas, lo cual es lógico, ya que las reacciones negativas no suelen estar directamente relacionadas con las positivas.

En resumen, los datos muestran que las publicaciones con más interacciones suelen seguir patrones similares (más compartidos y más reacciones positivas), mientras que las reacciones negativas no siguen las mismas tendencias. Esto podría ser útil para identificar qué tipo de contenido genera más interacciones o qué tipo de posts deberían evitarse para mejorar la percepción pública.