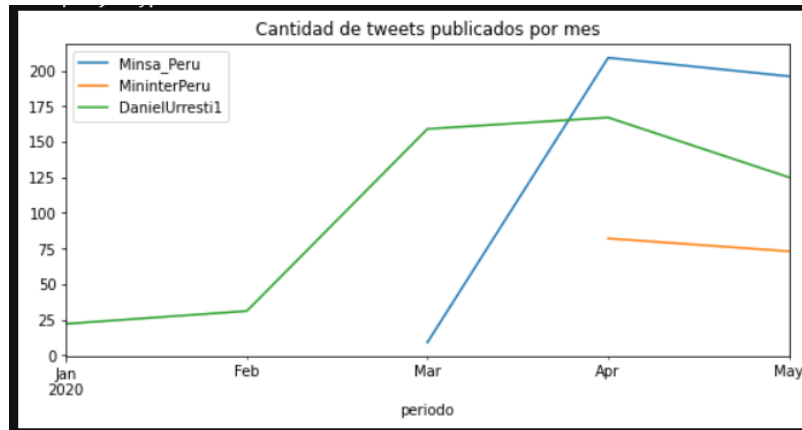


### Practica Nro 3

Víctor Mariano Villacorta Plasencia

#### 1- Análisis exploratorio

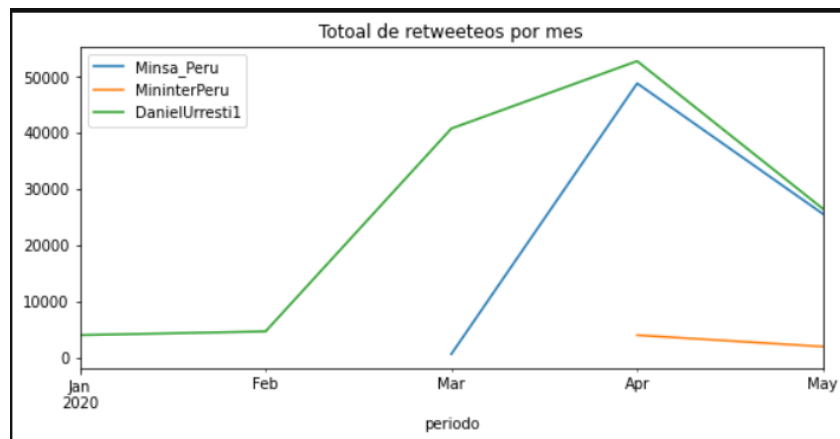


A simple inspección nuestro dataset cuenta con 3 cuentas. “Minsa\_Peru”, “MininterPeru” y “DanielUrresti”.

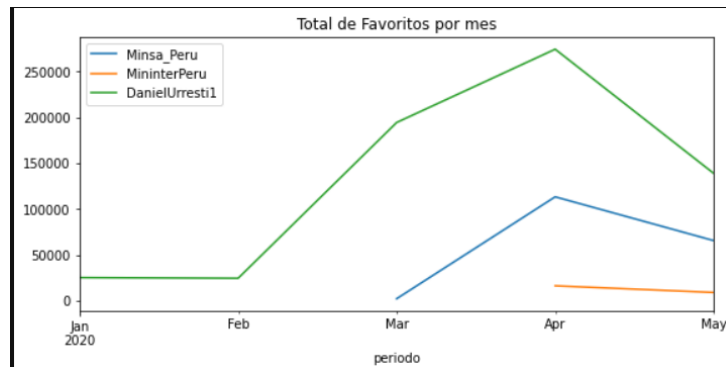
El rango total de información de publicación es desde Enero a mayo del 2020, de donde la cuenta de Minsa\_Peru es la que llega a mostrar mayor creación de tweets en los últimos meses pero registra los mismos desde Marzo.

Por otro lado la Cuenta de Daniel Urresti contiene información completa en todo el rango coincidiendo en mes pico con la cuenta del Minsa, pero aprecia mayor caída de creación de los mismos en el último mes.

Finalmente la cuenta de MiniterPeru, no muestra un comportamiento relevante en el periodo de evaluación donde se puede intuir que no se ve reflejado por los efectos coyunturales de las otras 2 cuentas.



A nivel de retweeteos totales por mes se aprecia una captura de la atención del público muy similar en las cuentas de Minsa\_Peru y DanielUrresti; y nuevamente se aprecia a la cuenta de Mininterperu fuera de contexto.



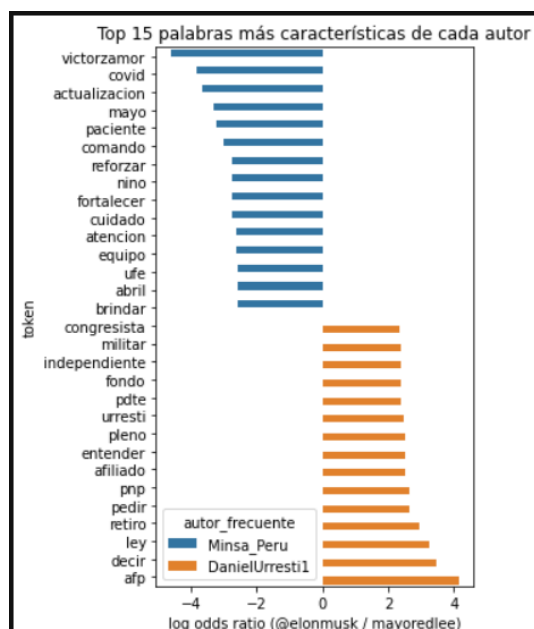
Para el caso de total de tweets marcados como favoritos se aprecia la misma tendencia en las cuentas de Minsa\_peru y DanielUrresti, pero un mayor valor para Daniel Urresti lo que resalta el respaldo político frente a las publicaciones de salud del Minsa.

	favoriteCount	retweetCount	n_words_cleaned
favoriteCount	1.000	0.968	0.103
retweetCount	0.968	1.000	0.106
n_words_cleaned	0.103	0.106	1.000

Agregando una función de correlación entre el conteo de favoritos, el conteo del retweets y el número de palabras por tweet(limpiado y lematizado), podemos ver, como ya sospechábamos que el conteo de los retweeteos tiene relación directa con el conteo de favoritos.

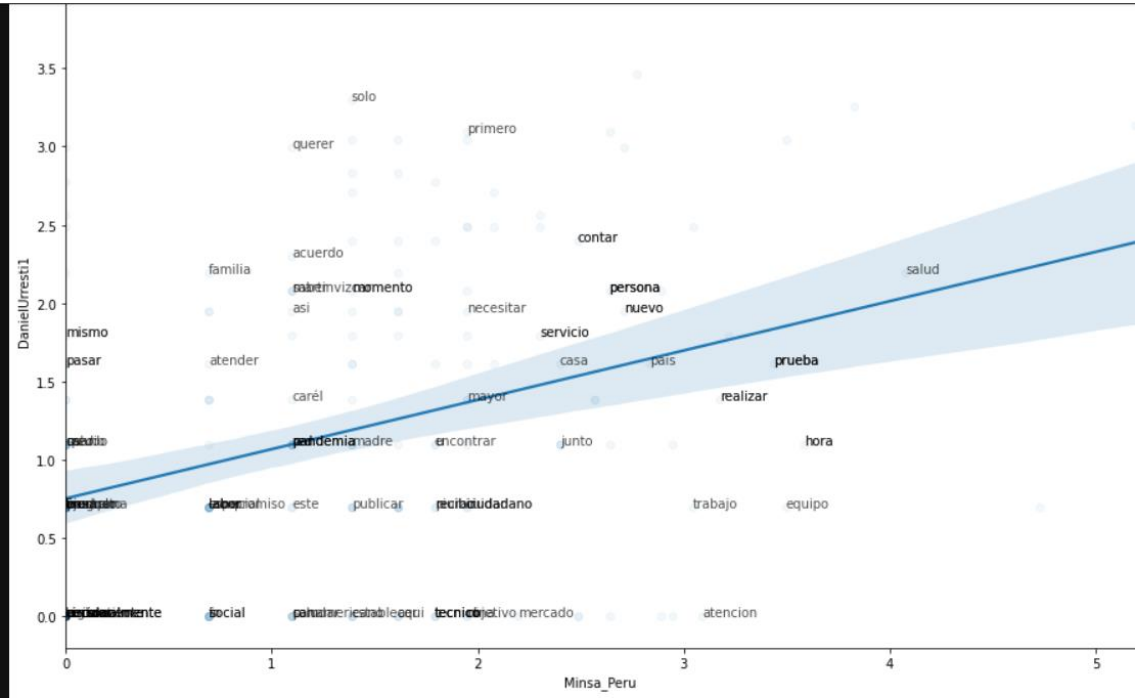
Por otro lado se descarta por completo que el número de palabras usadas por tweet pueda tener algún efectos sobre estos indicadores.

- Apartir de este punto me centraré en el análisis de las 2 cuentas mas relevantes



En la gráfica anterior apreciar que las palabras mas relevantes entre ambas cuentas son diferentes entre sí, lo cual me lleva a preguntarme.

¿Qué tan correlacionados estarán las cuentas?



De esta gráfica podemos ver que la pendiente no es tan pronuncia por lo cual nos lleva a la intuición que no son muy correlacionados.

	DanielUrresti1	MininterPeru	Minsa_Peru
DanielUrresti1	1.000	0.439	0.384
MininterPeru	0.439	1.000	0.755
Minsa_Peru	0.384	0.755	1.000

Para estar mas seguros se aplicó una diferencia de cosenos de donde ya se puede apreciar que con un valor de 0.384 no están muy correlacionados.

Lo cual puede ser muy útil al momento de entrenar algún modelo de predicción para diferencialo

