Download dataset di <https://github.com/ajihsan/sharing_agit/blob/master/Dataset.zip>

Load dataframe dari 3 data yang berada di folder exercise:

1. AkunTwitter\_POS.csv
2. HashtagTwitter\_POS.csv
3. Instagram\_POS.json

3 Data tersebut merupakan data sosial media dari PT. POS. Check schema dari masing-masing dataframe. Data tersebut berasal dari 2 sosial media yang berbeda: twitter & instagram.

Filter data yang berasal dari official account PT. POS di masing-masing sosial media.

Gabungan 3 dataframe diatas agar menjadi 1 format. Formatnya adalah:

**| Username | Content | Source |**

| bambang | Hi, posindo | Twitter |

| fathur12 | Halo, tolong | Instagram |

Hint: sesuaikan kolom 'content' agar merepresentasikan aspirasi user di masing-masing sosial media

Referensi:

<https://stackoverflow.com/questions/37471346/automatically-and-elegantly-flatten-dataframe-in-spark-sql><https://stackoverflow.com/questions/41027315/pyspark-split-multiple-array-columns-into-rows>