

Sudan Crisis

July 6, 2025

Human Rights Monitoring and Protection of Population in Sudan Using Twitter Data

```
[1]: !java -version
```

```
openjdk version "11.0.27" 2025-04-15
OpenJDK Runtime Environment (build 11.0.27+6-post-Ubuntu-0ubuntu122.04)
OpenJDK 64-Bit Server VM (build 11.0.27+6-post-Ubuntu-0ubuntu122.04, mixed mode,
sharing)
```

```
[2]: %pip install kafka-python pandas geopandas matplotlib seaborn numpy nltk spacy
↳ torch transformers pyspark findspark scikit-learn shapely tensorflow geopy
```

```
Requirement already satisfied: kafka-python in
/home/hadoop/py311env/lib/python3.11/site-packages (2.2.10)
Requirement already satisfied: pandas in
/home/hadoop/py311env/lib/python3.11/site-packages (2.2.3)
Requirement already satisfied: geopandas in
/home/hadoop/py311env/lib/python3.11/site-packages (1.0.1)
Requirement already satisfied: matplotlib in
/home/hadoop/py311env/lib/python3.11/site-packages (3.10.3)
Requirement already satisfied: seaborn in
/home/hadoop/py311env/lib/python3.11/site-packages (0.13.2)
Requirement already satisfied: numpy in
/home/hadoop/py311env/lib/python3.11/site-packages (2.1.3)
Requirement already satisfied: nltk in
/home/hadoop/py311env/lib/python3.11/site-packages (3.9.1)
Requirement already satisfied: spacy in
/home/hadoop/py311env/lib/python3.11/site-packages (3.8.7)
Requirement already satisfied: torch in
/home/hadoop/py311env/lib/python3.11/site-packages (2.7.0)
Requirement already satisfied: transformers in
/home/hadoop/py311env/lib/python3.11/site-packages (4.52.4)
Requirement already satisfied: pyspark in
/home/hadoop/py311env/lib/python3.11/site-packages (4.0.0)
Requirement already satisfied: findspark in
/home/hadoop/py311env/lib/python3.11/site-packages (2.0.1)
Requirement already satisfied: scikit-learn in
/home/hadoop/py311env/lib/python3.11/site-packages (1.6.1)
Requirement already satisfied: shapely in
```

/home/hadoop/py311env/lib/python3.11/site-packages (2.1.1)
 Requirement already satisfied: tensorflow in
 /home/hadoop/py311env/lib/python3.11/site-packages (2.19.0)
 Requirement already satisfied: geopy in
 /home/hadoop/py311env/lib/python3.11/site-packages (2.4.1)
 Requirement already satisfied: python-dateutil>=2.8.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from pandas) (2.9.0.post0)
 Requirement already satisfied: pytz>=2020.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from pandas) (2025.2)
 Requirement already satisfied: tzdata>=2022.7 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from pandas) (2025.2)
 Requirement already satisfied: pyogrio>=0.7.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from geopandas) (0.11.0)
 Requirement already satisfied: packaging in
 /home/hadoop/py311env/lib/python3.11/site-packages (from geopandas) (25.0)
 Requirement already satisfied: pyproj>=3.3.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from geopandas) (3.7.1)
 Requirement already satisfied: contourpy>=1.0.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (1.3.2)
 Requirement already satisfied: cyclor>=0.10 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (0.12.1)
 Requirement already satisfied: fonttools>=4.22.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (4.58.1)
 Requirement already satisfied: kiwisolver>=1.3.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (1.4.8)
 Requirement already satisfied: pillow>=8 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (11.2.1)
 Requirement already satisfied: pyparsing>=2.3.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (3.2.3)
 Requirement already satisfied: click in
 /home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (8.2.1)
 Requirement already satisfied: joblib in
 /home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (1.5.1)
 Requirement already satisfied: regex>=2021.8.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (2024.11.6)
 Requirement already satisfied: tqdm in
 /home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (4.67.1)
 Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.0.12)
 Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (1.0.5)
 Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (1.0.13)
 Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.0.11)
 Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.0.10)
 Requirement already satisfied: thinc<8.4.0,>=8.3.4 in

/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (8.3.6)
 Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (1.1.3)
 Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.5.1)
 Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.0.10)
 Requirement already satisfied: weasel<0.5.0,>=0.1.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (0.4.1)
 Requirement already satisfied: typer<1.0.0,>=0.3.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (0.16.0)
 Requirement already satisfied: requests<3.0.0,>=2.13.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.32.3)
 Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.11.5)
 Requirement already satisfied: jinja2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.1.6)
 Requirement already satisfied: setuptools in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (65.5.0)
 Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.5.0)
 Requirement already satisfied: language-data>=1.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 langcodes<4.0.0,>=3.2.0->spacy) (1.3.0)
 Requirement already satisfied: annotated-types>=0.6.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)
 Requirement already satisfied: pydantic-core==2.33.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (2.33.2)
 Requirement already satisfied: typing-extensions>=4.12.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (4.13.2)
 Requirement already satisfied: typing-inspection>=0.4.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.4.1)
 Requirement already satisfied: charset-normalizer<4,>=2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3.0.0,>=2.13.0->spacy) (3.4.2)
 Requirement already satisfied: idna<4,>=2.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3.0.0,>=2.13.0->spacy) (3.10)
 Requirement already satisfied: urllib3<3,>=1.21.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3.0.0,>=2.13.0->spacy) (2.4.0)
 Requirement already satisfied: certifi>=2017.4.17 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3.0.0,>=2.13.0->spacy) (2025.4.26)

Requirement already satisfied: blis<1.4.0,>=1.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
thinc<8.4.0,>=8.3.4->spacy) (1.3.0)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
thinc<8.4.0,>=8.3.4->spacy) (0.1.5)

Requirement already satisfied: shellingham>=1.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (1.5.4)

Requirement already satisfied: rich>=10.11.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (14.0.0)

Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (0.21.1)

Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (7.1.0)

Requirement already satisfied: wrapt in
/home/hadoop/py311env/lib/python3.11/site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy) (1.17.2)

Requirement already satisfied: filelock in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (3.18.0)

Requirement already satisfied: sympy>=1.13.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (1.14.0)

Requirement already satisfied: networkx in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (3.5)

Requirement already satisfied: fsspec in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (2025.5.1)

Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.77)

Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.77)

Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.80)

Requirement already satisfied: nvidia-cudnn-cu12==9.5.1.17 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (9.5.1.17)

Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.4.1)

Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (11.3.0.4)

Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (10.3.7.77)

Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (11.7.1.2)

Requirement already satisfied: nvidia-cuspars-cu12==12.5.4.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.5.4.2)

Requirement already satisfied: nvidia-cusparselt-cu12==0.6.3 in

```

/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (0.6.3)
Requirement already satisfied: nvidia-nccl-cu12==2.26.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (2.26.2)
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.85)
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (1.11.1.6)
Requirement already satisfied: triton==3.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (3.3.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.32.3)
Requirement already satisfied: pyyaml>=5.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (6.0.2)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.5.3)
Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (1.1.2)
Requirement already satisfied: py4j==0.10.9.9 in
/home/hadoop/py311env/lib/python3.11/site-packages (from pyspark) (0.10.9.9)
Requirement already satisfied: scipy>=1.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from scikit-learn) (1.15.3)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: absl-py>=1.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (2.3.0)
Requirement already satisfied: astunparse>=1.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers>=24.3.25 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (25.2.10)
Requirement already satisfied: gast!=0.5.0,!0.5.1,!0.5.2,>=0.2.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (0.6.0)
Requirement already satisfied: google-pasta>=0.1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (0.2.0)
Requirement already satisfied: libclang>=13.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (18.1.1)
Requirement already satisfied: opt-einsum>=2.3.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (3.4.0)
Requirement already satisfied:
protobuf!=4.21.0,!4.21.1,!4.21.2,!4.21.3,!4.21.4,!4.21.5,<6.0.0dev,>=3.20.3
in /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (5.29.5)
Requirement already satisfied: six>=1.12.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (1.17.0)
Requirement already satisfied: termcolor>=1.1.0 in

```

/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (3.1.0)
 Requirement already satisfied: grpcio<2.0,>=1.24.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (1.71.0)
 Requirement already satisfied: tensorboard~=2.19.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (2.19.0)
 Requirement already satisfied: keras>=3.5.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (3.10.0)
 Requirement already satisfied: h5py>=3.11.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (3.13.0)
 Requirement already satisfied: ml-dtypes<1.0.0,>=0.5.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (0.5.1)
 Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow) (0.37.1)
 Requirement already satisfied: markdown>=2.6.8 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 tensorboard~=2.19.0->tensorflow) (3.8)
 Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 tensorboard~=2.19.0->tensorflow) (0.7.2)
 Requirement already satisfied: werkzeug>=1.0.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 tensorboard~=2.19.0->tensorflow) (3.1.3)
 Requirement already satisfied: geographiclib<3,>=1.52 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from geopy) (2.0)
 Requirement already satisfied: wheel<1.0,>=0.23.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 astunparse>=1.6.0->tensorflow) (0.45.1)
 Requirement already satisfied: namex in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 keras>=3.5.0->tensorflow) (0.1.0)
 Requirement already satisfied: optree in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 keras>=3.5.0->tensorflow) (0.16.0)
 Requirement already satisfied: marisa-trie>=1.1.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from language-
 data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.1)
 Requirement already satisfied: markdown-it-py>=2.2.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (3.0.0)
 Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.19.1)
 Requirement already satisfied: mdurl~=0.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from markdown-it-
 py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)
 Requirement already satisfied: mpmath<1.4,>=1.1.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from sympy>=1.13.3->torch)
 (1.3.0)

Requirement already satisfied: MarkupSafe>=2.1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
werkzeug>=1.0.1->tensorboard~=2.19.0->tensorflow) (3.0.2)
Note: you may need to restart the kernel to use updated packages.

```
[3]: %pip install --upgrade jupyter ipywidgets
```

Requirement already satisfied: jupyter in
/home/hadoop/py311env/lib/python3.11/site-packages (1.1.1)
Requirement already satisfied: ipywidgets in
/home/hadoop/py311env/lib/python3.11/site-packages (8.1.7)
Requirement already satisfied: notebook in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (7.4.3)
Requirement already satisfied: jupyter-console in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (6.6.3)
Requirement already satisfied: nbconvert in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (7.16.6)
Requirement already satisfied: ipykernel in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (6.29.5)
Requirement already satisfied: jupyterlab in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (4.4.3)
Requirement already satisfied: comm>=0.1.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (0.2.2)
Requirement already satisfied: ipython>=6.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (9.2.0)
Requirement already satisfied: traitlets>=4.3.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (5.14.3)
Requirement already satisfied: widgetsnbextension~=4.0.14 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (4.0.14)
Requirement already satisfied: jupyterlab_widgets~=3.0.15 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (3.0.15)
Requirement already satisfied: decorator in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (5.2.1)
Requirement already satisfied: ipython-pygments-lexers in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (1.1.1)
Requirement already satisfied: jedi>=0.16 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (0.19.2)
Requirement already satisfied: matplotlib-inline in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (0.1.7)
Requirement already satisfied: pexpect>4.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (4.9.0)
Requirement already satisfied: prompt_toolkit<3.1.0,>=3.0.41 in
/home/hadoop/py311env/lib/python3.11/site-packages (from

```

ipython>=6.1.0->ipywidgets) (3.0.51)
Requirement already satisfied: pygments>=2.4.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (2.19.1)
Requirement already satisfied: stack_data in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (0.6.3)
Requirement already satisfied: typing_extensions>=4.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
ipython>=6.1.0->ipywidgets) (4.13.2)
Requirement already satisfied: wcwidth in
/home/hadoop/py311env/lib/python3.11/site-packages (from
prompt_toolkit<3.1.0,>=3.0.41->ipython>=6.1.0->ipywidgets) (0.2.13)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jedi>=0.16->ipython>=6.1.0->ipywidgets) (0.8.4)
Requirement already satisfied: ptyprocess>=0.5 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
pexpect>4.3->ipython>=6.1.0->ipywidgets) (0.7.0)
Requirement already satisfied: debugpy>=1.6.5 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(1.8.14)
Requirement already satisfied: jupyter-client>=6.1.12 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(8.6.3)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(5.8.1)
Requirement already satisfied: nest-asyncio in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(1.6.0)
Requirement already satisfied: packaging in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(25.0)
Requirement already satisfied: psutil in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(7.0.0)
Requirement already satisfied: pyzmq>=24 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(26.4.0)
Requirement already satisfied: tornado>=6.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
(6.5.1)
Requirement already satisfied: python-dateutil>=2.8.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
client>=6.1.12->ipykernel->jupyter) (2.9.0.post0)
Requirement already satisfied: platformdirs>=2.5 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-

```


core!=5.0.*,>=4.12->ipykernel->jupyter) (4.3.8)
 Requirement already satisfied: six>=1.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from python-
 dateutil>=2.8.2->jupyter-client>=6.1.12->ipykernel->jupyter) (1.17.0)
 Requirement already satisfied: async-lru>=1.0.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (2.0.5)
 Requirement already satisfied: httpx>=0.25.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (0.28.1)
 Requirement already satisfied: jinja2>=3.0.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (3.1.6)
 Requirement already satisfied: jupyter-lsp>=2.0.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (2.2.5)
 Requirement already satisfied: jupyter-server<3,>=2.4.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (2.16.0)
 Requirement already satisfied: jupyterlab-server<3,>=2.27.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (2.27.3)
 Requirement already satisfied: notebook-shim>=0.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (0.2.4)
 Requirement already satisfied: setuptools>=41.1.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (65.5.0)
 Requirement already satisfied: anyio>=3.1.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 server<3,>=2.4.0->jupyterlab->jupyter) (4.9.0)
 Requirement already satisfied: argon2-cffi>=21.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 server<3,>=2.4.0->jupyterlab->jupyter) (23.1.0)
 Requirement already satisfied: jupyter-events>=0.11.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 server<3,>=2.4.0->jupyterlab->jupyter) (0.12.0)
 Requirement already satisfied: jupyter-server-terminals>=0.4.4 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 server<3,>=2.4.0->jupyterlab->jupyter) (0.5.3)
 Requirement already satisfied: nbformat>=5.3.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 server<3,>=2.4.0->jupyterlab->jupyter) (5.10.4)
 Requirement already satisfied: overrides>=5.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 server<3,>=2.4.0->jupyterlab->jupyter) (7.7.0)
 Requirement already satisfied: prometheus-client>=0.9 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-

```

server<3,>=2.4.0->jupyterlab->jupyter) (0.22.0)
Requirement already satisfied: send2trash>=1.8.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.8.3)
Requirement already satisfied: terminado>=0.8.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (0.18.1)
Requirement already satisfied: websocket-client>=1.7 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.8.0)
Requirement already satisfied: babel>=2.10 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (2.17.0)
Requirement already satisfied: json5>=0.9.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (0.12.0)
Requirement already satisfied: jsonschema>=4.18.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (4.24.0)
Requirement already satisfied: requests>=2.31 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (2.32.3)
Requirement already satisfied: idna>=2.8 in
/home/hadoop/py311env/lib/python3.11/site-packages (from anyio>=3.1.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (3.10)
Requirement already satisfied: sniffio>=1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from anyio>=3.1.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.3.1)
Requirement already satisfied: argon2-cffi-bindings in
/home/hadoop/py311env/lib/python3.11/site-packages (from
argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (21.2.0)
Requirement already satisfied: certifi in
/home/hadoop/py311env/lib/python3.11/site-packages (from
httpx>=0.25.0->jupyterlab->jupyter) (2025.4.26)
Requirement already satisfied: httpcore==1.* in
/home/hadoop/py311env/lib/python3.11/site-packages (from
httpx>=0.25.0->jupyterlab->jupyter) (1.0.9)
Requirement already satisfied: h11>=0.16 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
httpcore==1.*->httpx>=0.25.0->jupyterlab->jupyter) (0.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jinja2>=3.0.3->jupyterlab->jupyter) (3.0.2)
Requirement already satisfied: attrs>=22.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from

```

```

jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter)
(2025.4.1)
Requirement already satisfied: referencing>=0.28.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (0.25.1)
Requirement already satisfied: python-json-logger>=2.0.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (3.3.0)
Requirement already satisfied: pyyaml>=5.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (6.0.2)
Requirement already satisfied: rfc3339-validator in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (0.1.4)
Requirement already satisfied: rfc3986-validator>=0.1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (0.1.1)
Requirement already satisfied: fqdn in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.5.1)
Requirement already satisfied: isoduration in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (20.11.0)
Requirement already satisfied: jsonpointer>1.13 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (3.0.0)
Requirement already satisfied: uri-template in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.3.0)
Requirement already satisfied: webcolors>=24.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (24.11.1)
Requirement already satisfied: beautifulsoup4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(4.13.4)
Requirement already satisfied: bleach!=5.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
bleach[css]!=5.0.0->nbconvert->jupyter) (6.2.0)
Requirement already satisfied: defusedxml in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)

```

(0.7.1)
Requirement already satisfied: jupyterlab-pygments in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(0.3.0)
Requirement already satisfied: mistune<4,>=2.0.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(3.1.3)
Requirement already satisfied: nbclient>=0.5.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(0.10.2)
Requirement already satisfied: pandocfilters>=1.4.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(1.5.1)
Requirement already satisfied: webencodings in
/home/hadoop/py311env/lib/python3.11/site-packages (from
bleach!=5.0.0->bleach[css]!=5.0.0->nbconvert->jupyter) (0.5.1)
Requirement already satisfied: tinycss2<1.5,>=1.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
bleach[css]!=5.0.0->nbconvert->jupyter) (1.4.0)
Requirement already satisfied: fastjsonschema>=2.15 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
nbformat>=5.3.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (2.21.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
requests>=2.31->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (3.4.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
requests>=2.31->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (2.4.0)
Requirement already satisfied: cffi>=1.0.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from argon2-cffi-
bindings->argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
(1.17.1)
Requirement already satisfied: pycparser in
/home/hadoop/py311env/lib/python3.11/site-packages (from
cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (2.22)
Requirement already satisfied: soupsieve>1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
beautifulsoup4->nbconvert->jupyter) (2.7)
Requirement already satisfied: arrow>=0.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.3.0)
Requirement already satisfied: types-python-dateutil>=2.8.10 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
arrow>=0.15.0->isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (2.9.0.20250516)
Requirement already satisfied: executing>=1.2.0 in

```

/home/hadoop/py311env/lib/python3.11/site-packages (from
stack_data->ipython>=6.1.0->ipywidgets) (2.2.0)
Requirement already satisfied: asttokens>=2.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
stack_data->ipython>=6.1.0->ipywidgets) (3.0.0)
Requirement already satisfied: pure-eval in
/home/hadoop/py311env/lib/python3.11/site-packages (from
stack_data->ipython>=6.1.0->ipywidgets) (0.2.3)
Note: you may need to restart the kernel to use updated packages.

```

```
[4]: !jupyter nbextension enable --py widgetsnbextension
```

```

usage: jupyter [-h] [--version] [--config-dir] [--data-dir] [--runtime-dir]
              [--paths] [--json] [--debug]
              [subcommand]

```

Jupyter: Interactive Computing

positional arguments:

subcommand the subcommand to launch

options:

```

-h, --help      show this help message and exit
--version       show the versions of core jupyter packages and exit
--config-dir    show Jupyter config dir
--data-dir      show Jupyter data dir
--runtime-dir   show Jupyter runtime dir
--paths         show all Jupyter paths. Add --json for machine-readable
                format.
--json          output paths as machine-readable json
--debug         output debug information about paths

```

Available subcommands: console dejavu events execute kernel kernelspec lab
labextension labhub migrate nbconvert notebook run server troubleshoot trust

Jupyter command `jupyter-nbextension` not found.

```
[5]: import sys
print(sys.executable)
```

```
/home/hadoop/py311env/bin/python
```

```
[6]: import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-11-openjdk-amd64"
os.environ["SPARK_HOME"] = "/opt/spark/spark-3.5.6-bin-hadoop3"
os.environ["HADOOP_HOME"] = "/usr/local/hadoop"
```

```
[7]: import findspark
findspark.init()
import pyspark
```

```
[8]: #Import necessary libraries
import pandas as pd    # Pandas is an open source, BSD-licensed library
    ↳ providing high-performance, easy-to-use data structures and data analysis
    ↳ tools for the Python programming language. Pandas Series is a
    ↳ one-dimensional labeled array capable of holding any data type with axis
    ↳ labels or index
import geopandas as gpd
import time
import matplotlib.pyplot as plt
import numpy as np
import logging
import hashlib
from geopy.geocoders import Nominatim
from geopy.extra.rate_limiter import RateLimiter
from pyspark.sql import functions as func
from pyspark.sql.types import IntegerType, StringType, FloatType,
    ↳ TimestampType, BinaryType
from pyspark.sql.functions import col, count, when, mean, stddev, abs, sum,
    ↳ row_number
from pyspark.ml.feature import VectorAssembler, StandardScaler, StringIndexer,
    ↳ OneHotEncoder
from pyspark.ml import Pipeline
from pyspark.ml.classification import RandomForestClassifier
from pyspark.ml.tuning import CrossValidator, ParamGridBuilder
from pyspark.ml.evaluation import MulticlassClassificationEvaluator,
    ↳ BinaryClassificationEvaluator
from sklearn.metrics import precision_recall_curve
from pyspark.sql import Window
from transformers import pipeline, AutoTokenizer,
    ↳ AutoModelForSequenceClassification
from shapely.geometry import Point
import torch
```

2025-07-05 05:22:48.313695: E

external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:467] Unable to register cuFFT factory: Attempting to register factory for plugin cuFFT when one has already been registered

WARNING: All log messages before absl::InitializeLog() is called are written to STDERR

E0000 00:00:1751685768.579207 4919 cuda_dnn.cc:8579] Unable to register cuDNN factory: Attempting to register factory for plugin cuDNN when one has already been registered

E0000 00:00:1751685768.654715 4919 cuda_blas.cc:1407] Unable to register

```
cuBLAS factory: Attempting to register factory for plugin cuBLAS when one has
already been registered
W0000 00:00:1751685769.248405    4919 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
W0000 00:00:1751685769.248476    4919 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
W0000 00:00:1751685769.248478    4919 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
W0000 00:00:1751685769.248479    4919 computation_placer.cc:177] computation
placer already registered. Please check linkage and avoid linking the same
target more than once.
2025-07-05 05:22:49.319856: I tensorflow/core/platform/cpu_feature_guard.cc:210]
This TensorFlow binary is optimized to use available CPU instructions in
performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations, rebuild
TensorFlow with the appropriate compiler flags.
```

```
[9]: import nltk
nltk.download('punkt')    # Download NLTK essential resources
!python -m spacy download en_core_web_sm    # Download spaCy Language Model
```

```
[nltk_data] Downloading package punkt to /home/hadoop/nltk_data...
```

```
[nltk_data] Package punkt is already up-to-date!
```

```
Collecting en-core-web-sm==3.8.0
```

```
  Downloading https://github.com/explosion/spacy-
models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-
any.whl (12.8 MB)
```

```
12.8/12.8 MB
```

```
32.5 MB/s eta 0:00:0000:010:01
```

```
Download and installation successful
```

```
You can now load the package via spacy.load('en_core_web_sm')
```

```
[10]: import nltk
nltk.download('punkt_tab')
```

```
[nltk_data] Downloading package punkt_tab to /home/hadoop/nltk_data...
```

```
[nltk_data] Package punkt_tab is already up-to-date!
```

```
[10]: True
```

```
[11]: # Test NLTK installation
from nltk.tokenize import word_tokenize
text = "Hello, world!"
tokens = word_tokenize(text)
```

```
print(tokens)
```

```
['Hello', ',', 'world', '!']
```

```
[12]: import spacy
      try:
          nlp = spacy.load("en_core_web_sm")
      except OSError:
          from spacy.cli import download
          download("en_core_web_sm")
          nlp = spacy.load("en_core_web_sm")

      # Test spaCy installation
      doc = nlp("Hello, world!")
      print([token.text for token in doc])
```

```
['Hello', ',', 'world', '!']
```

```
[13]: # Start SparkSession
      from pyspark.sql import SparkSession
      spark=SparkSession.builder.appName("Sudan Crisis Analysis").config("spark.
      ↪driver.memory", "4g").getOrCreate()
      spark
```

```
25/07/05 05:23:15 WARN Utils: Your hostname, julius-virtual-machine resolves to
a loopback address: 127.0.1.1; using 192.168.145.130 instead (on interface
ens33)
```

```
25/07/05 05:23:15 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
```

```
Setting default log level to "WARN".
```

```
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLevel).
```

```
25/07/05 05:23:16 WARN NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
```

```
[13]: <pyspark.sql.session.SparkSession at 0x7d7b6b8e7950>
```

```
[ ]:
```


1 Data Sourcing and Pre-processing

2 Buy 35,000 twitter pos for 160USD

3 <https://www.tweetbinder.com/>

```
[14]: #Load the data and handle missing values
df_pyspark = spark.read.csv("Sudan_Twitter_Data_June_2025.csv", header=True,
    inferSchema=True).cache()
df_pyspark.show(5)
```

25/07/05 05:23:33 WARN SparkStringUtils: Truncated the string representation of a plan since it was too large. This behavior can be adjusted by setting 'spark.sql.debug.maxToStringFields'.

[Stage 2:=====> (1 + 2) / 3]

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      _id|      userId|      userName|      alias|      profilePicture|
userCreatedAt|
bio|followers|following|lists|statusesCount|verified|userLocation|lang|  type|
text|      createdAt|      mentions|totalRetweets|favorites|replies|l
inks|firstLink|images|impressions|isImage|firstImage| location|
url|noteTweetText|inReplyTo|inReplyToId|inReplyToUserId|reportName|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|1.93482E+18|2.43752092E8|      Damascene|
nanssolo|https://pbs.twimg...|2011-01-27T18:51:...|Syrian* blunt and...|
941|      754|      2|      182403|      FALSE|      NULL|      en|retweet|RT
@BSonblast: Pl...|2025-06-17T03:32:...|      "["@bsonblast"]"|      0|
0|      0|      []|      NULL|      []|      0|      FALSE|      NULL|Abyei
PCA|https://x.com/use...|      NULL|      FALSE|      NULL|      NULL|
NULL|
|1.93482E+18|      1.44527E18|      nana nano|
hydrowanter|https://pbs.twimg...|2021-10-05T06:07:...|she/her | Indones...|
156|      46|      1|      37124|      FALSE|      NULL|      en|retweet|RT
@BSonblast: Pl...|2025-06-17T03:31:...|      "["@bsonblast"]"|      0|
0|      0|      []|      NULL|      []|      0|      FALSE|      NULL|Abyei
PCA|https://x.com/use...|      NULL|      FALSE|      NULL|      NULL|
NULL|
```

```
|1.93482E+18| 1.06232E18| bean |
Charlixxx_|https://pbs.twimg...|2018-11-13T12:11:...| pan pirate ...|
152| 327| 0| 35414| FALSE| NULL| en|retweet|RT
@BSonblast: Pl...|2025-06-17T03:30:...| "["@bsonblast"]"| 0|
0| 0| []| NULL| []| 0| FALSE| NULL|Abyei
PCA|https://x.com/use...| NULL| FALSE| NULL| NULL|
NULL|
|1.93482E+18| 7.15708E17| Pastel
|soIguessIexist|https://pbs.twimg...|2016-04-01T01:12:...|24 | black | nonb...|
488| 554| 2| 36040| FALSE| NULL| en|retweet|RT
@blkdesertfair...|2025-06-17T03:30:...| "["@blkdesertfai...| 0|
0| 0| []| NULL| []| 0| FALSE| NULL|Abyei
PCA|https://x.com/use...| NULL| FALSE| NULL| NULL|
NULL|
|1.93482E+18| 1.2088E18| ui |
02A4D3|https://pbs.twimg...|2019-12-22T17:21:...| NULL| 14|
204| 0| 2549| FALSE| NULL| en|retweet|RT @BSonblast:
Pl...|2025-06-17T03:29:...| "["@bsonblast"]"| 0| 0| 0|
[]| NULL| []| 0| FALSE| NULL|Abyei
PCA|https://x.com/use...| NULL| FALSE| NULL| NULL|
NULL|
+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+
only showing top 5 rows
```

```
[15]: #View schema
df_pyspark.printSchema()
```

```
root
|-- _id: string (nullable = true)
|-- userId: double (nullable = true)
|-- userName: string (nullable = true)
|-- alias: string (nullable = true)
|-- profilePicture: string (nullable = true)
|-- userCreatedAt: string (nullable = true)
|-- bio: string (nullable = true)
|-- followers: string (nullable = true)
|-- following: string (nullable = true)
|-- lists: string (nullable = true)
|-- statusesCount: string (nullable = true)
|-- verified: string (nullable = true)
```

```

|-- userLocation: string (nullable = true)
|-- lang: string (nullable = true)
|-- type: string (nullable = true)
|-- text: string (nullable = true)
|-- createdAt: string (nullable = true)
|-- mentions: string (nullable = true)
|-- totalRetweets: string (nullable = true)
|-- favorites: string (nullable = true)
|-- replies: string (nullable = true)
|-- links: string (nullable = true)
|-- firstLink: string (nullable = true)
|-- images: string (nullable = true)
|-- impressions: string (nullable = true)
|-- isImage: string (nullable = true)
|-- firstImage: string (nullable = true)
|-- location: string (nullable = true)
|-- url: string (nullable = true)
|-- noteTweetText: string (nullable = true)
|-- inReplyTo: string (nullable = true)
|-- inReplyToId: string (nullable = true)
|-- inReplyToUserId: string (nullable = true)
|-- reportName: string (nullable = true)

```

```

[16]: # Convert PySpark DataFrame to Pandas
filtered_df_pandas = df_pyspark.toPandas()

```

```

[17]: from geopy.geocoders import Nominatim
from geopy.extra.rate_limiter import RateLimiter
import pandas as pd

geolocator = Nominatim(user_agent="geoapi", timeout=10)
geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1,
↳error_wait_seconds=1)

def get_lat_lon(location):
    if not location or len(str(location).strip()) < 3:
        return pd.Series([None, None])
    try:
        loc = geocode(location)
        if loc:
            return pd.Series([loc.latitude, loc.longitude])
        else:
            return pd.Series([None, None])
    except Exception as e:
        return pd.Series([None, None])

```

```
filtered_df_pandas[['Latitude', 'Longitude']] = filtered_df_pandas['location'].  
    ↪apply(get_lat_lon)
```

```
[18]: # Export dataframe  
filtered_df_pandas.to_csv('filtered_data.csv', index=False)
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```