

# Sample\_extraction\_code

July 10, 2025

```
[1]: # utilities
import re
import pickle
import numpy as np
import pandas as pd

# important libraries
from bokeh.plotting import figure
from bokeh.io import output_file, show, output_notebook
from collections import Counter
import spacy
from spacy.util import compounding
from spacy.util import minibatch
from spacy import displacy
import gc
import os
from scipy.stats import skew, kurtosis

# sklearn
from sklearn.svm import LinearSVC
from sklearn.naive_bayes import BernoulliNB
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier

from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import confusion_matrix, classification_report
```

```
[2]: # Set the folder path
user_path = r"E:\Data_Mining_project\Data mining data" #give file path of
↳Data_Mining_project

# Define what fraction of each file to sample
```

```

sample_fraction = 0.01 # 5% sample
random_state = 42      # for reproducibility

#Gather all CSV paths
csv_collection = []
for dirname, _, filenames in os.walk(user_path):
    for filename in filenames:
        if filename.endswith('.csv') or filename.endswith('.csv.gz'):
            csv_collection.append(os.path.join(dirname, filename))

print(f" Total CSV files found: {len(csv_collection)}")

#Load and sample from each file
sampled_chunks = []

for i, file in enumerate(csv_collection):
    try:
        df_chunk = pd.read_csv(file, index_col=0, low_memory=False)
        sample = df_chunk.sample(frac=sample_fraction,
    ↪random_state=random_state)
        sampled_chunks.append(sample)
        print(f" Sampled {len(sample)} rows from: {file}")
    except Exception as e:
        print(f" Error loading {file}: {e}")

#Concatenate all sampled parts
df = pd.concat(sampled_chunks, axis=0, ignore_index=True)
print(f"\n Final sampled DataFrame shape: {df.shape}")

```

```

Total CSV files found: 461
Sampled 3649 rows from: E:\Data_Mining_project\Data mining
data\0401_UkraineCombinedTweetsDeduped.csv
Sampled 3710 rows from: E:\Data_Mining_project\Data mining
data\0402_UkraineCombinedTweetsDeduped.csv
Sampled 4455 rows from: E:\Data_Mining_project\Data mining
data\0403_UkraineCombinedTweetsDeduped.csv
Sampled 4304 rows from: E:\Data_Mining_project\Data mining
data\0404_UkraineCombinedTweetsDeduped.csv
Sampled 4526 rows from: E:\Data_Mining_project\Data mining
data\0405_UkraineCombinedTweetsDeduped.csv
Sampled 4228 rows from: E:\Data_Mining_project\Data mining
data\0406_UkraineCombinedTweetsDeduped.csv
Sampled 3787 rows from: E:\Data_Mining_project\Data mining
data\0407_UkraineCombinedTweetsDeduped.csv
Sampled 4055 rows from: E:\Data_Mining_project\Data mining
data\0408_UkraineCombinedTweetsDeduped.csv
Sampled 4030 rows from: E:\Data_Mining_project\Data mining
data\0409_UkraineCombinedTweetsDeduped.csv

```

Sampled 3731 rows from: E:\Data\_Mining\_project\Data mining data\0410\_UkraineCombinedTweetsDeduped.csv  
Sampled 3762 rows from: E:\Data\_Mining\_project\Data mining data\0411\_UkraineCombinedTweetsDeduped.csv  
Sampled 3953 rows from: E:\Data\_Mining\_project\Data mining data\0412\_UkraineCombinedTweetsDeduped.csv  
Sampled 3883 rows from: E:\Data\_Mining\_project\Data mining data\0413\_UkraineCombinedTweetsDeduped.csv  
Sampled 3666 rows from: E:\Data\_Mining\_project\Data mining data\0414\_UkraineCombinedTweetsDeduped.csv  
Sampled 3580 rows from: E:\Data\_Mining\_project\Data mining data\0415\_UkraineCombinedTweetsDeduped.csv  
Sampled 3329 rows from: E:\Data\_Mining\_project\Data mining data\0416\_UkraineCombinedTweetsDeduped.csv  
Sampled 3259 rows from: E:\Data\_Mining\_project\Data mining data\0417\_UkraineCombinedTweetsDeduped.csv  
Sampled 3359 rows from: E:\Data\_Mining\_project\Data mining data\0418\_UkraineCombinedTweetsDeduped.csv  
Sampled 4165 rows from: E:\Data\_Mining\_project\Data mining data\0419\_UkraineCombinedTweetsDeduped.csv  
Sampled 3972 rows from: E:\Data\_Mining\_project\Data mining data\0420\_UkraineCombinedTweetsDeduped.csv  
Sampled 3612 rows from: E:\Data\_Mining\_project\Data mining data\0421\_UkraineCombinedTweetsDeduped.csv  
Sampled 3438 rows from: E:\Data\_Mining\_project\Data mining data\0422\_UkraineCombinedTweetsDeduped.csv  
Sampled 3620 rows from: E:\Data\_Mining\_project\Data mining data\0423\_UkraineCombinedTweetsDeduped.csv  
Sampled 3490 rows from: E:\Data\_Mining\_project\Data mining data\0424\_UkraineCombinedTweetsDeduped.csv  
Sampled 3177 rows from: E:\Data\_Mining\_project\Data mining data\0425\_UkraineCombinedTweetsDeduped.csv  
Sampled 3434 rows from: E:\Data\_Mining\_project\Data mining data\0426\_UkraineCombinedTweetsDeduped.csv  
Sampled 3188 rows from: E:\Data\_Mining\_project\Data mining data\0427\_UkraineCombinedTweetsDeduped.csv  
Sampled 3152 rows from: E:\Data\_Mining\_project\Data mining data\0428\_UkraineCombinedTweetsDeduped.csv  
Sampled 3286 rows from: E:\Data\_Mining\_project\Data mining data\0429\_UkraineCombinedTweetsDeduped.csv  
Sampled 3281 rows from: E:\Data\_Mining\_project\Data mining data\0430\_UkraineCombinedTweetsDeduped.csv  
Sampled 3519 rows from: E:\Data\_Mining\_project\Data mining data\0501\_UkraineCombinedTweetsDeduped.csv  
Sampled 3295 rows from: E:\Data\_Mining\_project\Data mining data\0502\_UkraineCombinedTweetsDeduped.csv  
Sampled 3071 rows from: E:\Data\_Mining\_project\Data mining data\0503\_UkraineCombinedTweetsDeduped.csv

Sampled 3843 rows from: E:\Data\_Mining\_project\Data mining data\0504\_UkraineCombinedTweetsDeduped.csv  
Sampled 10462 rows from: E:\Data\_Mining\_project\Data mining data\0505\_to\_0507\_UkraineCombinedTweetsDeduped.csv  
Sampled 3717 rows from: E:\Data\_Mining\_project\Data mining data\0508\_UkraineCombinedTweetsDeduped.csv  
Sampled 3991 rows from: E:\Data\_Mining\_project\Data mining data\0509\_UkraineCombinedTweetsDeduped.csv  
Sampled 3526 rows from: E:\Data\_Mining\_project\Data mining data\0510\_UkraineCombinedTweetsDeduped.csv  
Sampled 3101 rows from: E:\Data\_Mining\_project\Data mining data\0511\_UkraineCombinedTweetsDeduped.csv  
Sampled 3025 rows from: E:\Data\_Mining\_project\Data mining data\0512\_UkraineCombinedTweetsDeduped.csv  
Sampled 2802 rows from: E:\Data\_Mining\_project\Data mining data\0513\_UkraineCombinedTweetsDeduped.csv  
Sampled 3262 rows from: E:\Data\_Mining\_project\Data mining data\0514\_UkraineCombinedTweetsDeduped.csv  
Sampled 4202 rows from: E:\Data\_Mining\_project\Data mining data\0515\_UkraineCombinedTweetsDeduped.csv  
Sampled 3879 rows from: E:\Data\_Mining\_project\Data mining data\0516\_UkraineCombinedTweetsDeduped.csv  
Sampled 3379 rows from: E:\Data\_Mining\_project\Data mining data\0517\_UkraineCombinedTweetsDeduped.csv  
Sampled 3192 rows from: E:\Data\_Mining\_project\Data mining data\0518\_UkraineCombinedTweetsDeduped.csv  
Sampled 2722 rows from: E:\Data\_Mining\_project\Data mining data\0519\_UkraineCombinedTweetsDeduped.csv  
Sampled 2298 rows from: E:\Data\_Mining\_project\Data mining data\0520\_UkraineCombinedTweetsDeduped.csv  
Sampled 2116 rows from: E:\Data\_Mining\_project\Data mining data\0521\_UkraineCombinedTweetsDeduped.csv  
Sampled 1979 rows from: E:\Data\_Mining\_project\Data mining data\0522\_UkraineCombinedTweetsDeduped.csv  
Sampled 2450 rows from: E:\Data\_Mining\_project\Data mining data\0523\_UkraineCombinedTweetsDeduped.csv  
Sampled 2466 rows from: E:\Data\_Mining\_project\Data mining data\0524\_UkraineCombinedTweetsDeduped.csv  
Sampled 2317 rows from: E:\Data\_Mining\_project\Data mining data\0525\_UkraineCombinedTweetsDeduped.csv  
Sampled 2185 rows from: E:\Data\_Mining\_project\Data mining data\0526\_UkraineCombinedTweetsDeduped.csv  
Sampled 2383 rows from: E:\Data\_Mining\_project\Data mining data\0527\_UkraineCombinedTweetsDeduped.csv  
Sampled 2296 rows from: E:\Data\_Mining\_project\Data mining data\0528\_UkraineCombinedTweetsDeduped.csv  
Sampled 2199 rows from: E:\Data\_Mining\_project\Data mining data\0529\_UkraineCombinedTweetsDeduped.csv

Sampled 2340 rows from: E:\Data\_Mining\_project\Data mining data\0530\_UkraineCombinedTweetsDeduped.csv  
Sampled 2179 rows from: E:\Data\_Mining\_project\Data mining data\0531\_UkraineCombinedTweetsDeduped.csv  
Sampled 2685 rows from: E:\Data\_Mining\_project\Data mining data\0601\_UkraineCombinedTweetsDeduped.csv  
Sampled 2144 rows from: E:\Data\_Mining\_project\Data mining data\0602\_UkraineCombinedTweetsDeduped.csv  
Sampled 2295 rows from: E:\Data\_Mining\_project\Data mining data\0603\_UkraineCombinedTweetsDeduped.csv  
Sampled 2381 rows from: E:\Data\_Mining\_project\Data mining data\0604\_UkraineCombinedTweetsDeduped.csv  
Sampled 2151 rows from: E:\Data\_Mining\_project\Data mining data\0605\_UkraineCombinedTweetsDeduped.csv  
Sampled 7058 rows from: E:\Data\_Mining\_project\Data mining data\0606\_to\_08\_UkraineCombinedTweetsDeduped.csv  
Sampled 2646 rows from: E:\Data\_Mining\_project\Data mining data\0609\_UkraineCombinedTweetsDeduped.csv  
Sampled 2592 rows from: E:\Data\_Mining\_project\Data mining data\0610\_UkraineCombinedTweetsDeduped.csv  
Sampled 2204 rows from: E:\Data\_Mining\_project\Data mining data\0611\_UkraineCombinedTweetsDeduped.csv  
Sampled 2164 rows from: E:\Data\_Mining\_project\Data mining data\0612\_UkraineCombinedTweetsDeduped.csv  
Sampled 2183 rows from: E:\Data\_Mining\_project\Data mining data\0613\_UkraineCombinedTweetsDeduped.csv  
Sampled 2185 rows from: E:\Data\_Mining\_project\Data mining data\0614\_UkraineCombinedTweetsDeduped.csv  
Sampled 2159 rows from: E:\Data\_Mining\_project\Data mining data\0615\_UkraineCombinedTweetsDeduped.csv  
Sampled 2888 rows from: E:\Data\_Mining\_project\Data mining data\0616\_UkraineCombinedTweetsDeduped.csv  
Sampled 2619 rows from: E:\Data\_Mining\_project\Data mining data\0617\_UkraineCombinedTweetsDeduped.csv  
Sampled 2052 rows from: E:\Data\_Mining\_project\Data mining data\0618\_UkraineCombinedTweetsDeduped.csv  
Sampled 1788 rows from: E:\Data\_Mining\_project\Data mining data\0619\_UkraineCombinedTweetsDeduped.csv  
Sampled 1882 rows from: E:\Data\_Mining\_project\Data mining data\0620\_UkraineCombinedTweetsDeduped.csv  
Sampled 1857 rows from: E:\Data\_Mining\_project\Data mining data\0621\_UkraineCombinedTweetsDeduped.csv  
Sampled 1960 rows from: E:\Data\_Mining\_project\Data mining data\0622\_UkraineCombinedTweetsDeduped.csv  
Sampled 2174 rows from: E:\Data\_Mining\_project\Data mining data\0623\_UkraineCombinedTweetsDeduped.csv  
Sampled 2039 rows from: E:\Data\_Mining\_project\Data mining data\0624\_UkraineCombinedTweetsDeduped.csv

Sampled 1737 rows from: E:\Data\_Mining\_project\Data mining data\0625\_UkraineCombinedTweetsDeduped.csv  
Sampled 2132 rows from: E:\Data\_Mining\_project\Data mining data\0626\_UkraineCombinedTweetsDeduped.csv  
Sampled 2711 rows from: E:\Data\_Mining\_project\Data mining data\0627\_UkraineCombinedTweetsDeduped.csv  
Sampled 2844 rows from: E:\Data\_Mining\_project\Data mining data\0628\_UkraineCombinedTweetsDeduped.csv  
Sampled 2822 rows from: E:\Data\_Mining\_project\Data mining data\0629\_UkraineCombinedTweetsDeduped.csv  
Sampled 2198 rows from: E:\Data\_Mining\_project\Data mining data\0630\_UkraineCombinedTweetsDeduped.csv  
Sampled 1952 rows from: E:\Data\_Mining\_project\Data mining data\0701\_UkraineCombinedTweetsDeduped.csv  
Sampled 1668 rows from: E:\Data\_Mining\_project\Data mining data\0702\_UkraineCombinedTweetsDeduped.csv  
Sampled 1557 rows from: E:\Data\_Mining\_project\Data mining data\0703\_UkraineCombinedTweetsDeduped.csv  
Sampled 1814 rows from: E:\Data\_Mining\_project\Data mining data\0704\_UkraineCombinedTweetsDeduped.csv  
Sampled 1715 rows from: E:\Data\_Mining\_project\Data mining data\0705\_UkraineCombinedTweetsDeduped.csv  
Sampled 1606 rows from: E:\Data\_Mining\_project\Data mining data\0706\_UkraineCombinedTweetsDeduped.csv  
Sampled 1721 rows from: E:\Data\_Mining\_project\Data mining data\0707\_UkraineCombinedTweetsDeduped.csv  
Sampled 1769 rows from: E:\Data\_Mining\_project\Data mining data\0708\_UkraineCombinedTweetsDeduped.csv  
Sampled 1569 rows from: E:\Data\_Mining\_project\Data mining data\0709\_UkraineCombinedTweetsDeduped.csv  
Sampled 1789 rows from: E:\Data\_Mining\_project\Data mining data\0710\_UkraineCombinedTweetsDeduped.csv  
Sampled 2311 rows from: E:\Data\_Mining\_project\Data mining data\0711\_UkraineCombinedTweetsDeduped.csv  
Sampled 1882 rows from: E:\Data\_Mining\_project\Data mining data\0712\_UkraineCombinedTweetsDeduped.csv  
Sampled 1641 rows from: E:\Data\_Mining\_project\Data mining data\0713\_UkraineCombinedTweetsDeduped.csv  
Sampled 2487 rows from: E:\Data\_Mining\_project\Data mining data\0714\_UkraineCombinedTweetsDeduped.csv  
Sampled 2185 rows from: E:\Data\_Mining\_project\Data mining data\0715\_UkraineCombinedTweetsDeduped.csv  
Sampled 1616 rows from: E:\Data\_Mining\_project\Data mining data\0716\_UkraineCombinedTweetsDeduped.csv  
Sampled 1586 rows from: E:\Data\_Mining\_project\Data mining data\0717\_UkraineCombinedTweetsDeduped.csv  
Sampled 1629 rows from: E:\Data\_Mining\_project\Data mining data\0718\_UkraineCombinedTweetsDeduped.csv

Sampled 1594 rows from: E:\Data\_Mining\_project\Data mining data\0719\_UkraineCombinedTweetsDeduped.csv  
Sampled 1730 rows from: E:\Data\_Mining\_project\Data mining data\0720\_UkraineCombinedTweetsDeduped.csv  
Sampled 1492 rows from: E:\Data\_Mining\_project\Data mining data\0721\_UkraineCombinedTweetsDeduped.csv  
Sampled 1351 rows from: E:\Data\_Mining\_project\Data mining data\0722\_UkraineCombinedTweetsDeduped.csv  
Sampled 1303 rows from: E:\Data\_Mining\_project\Data mining data\0723\_UkraineCombinedTweetsDeduped.csv  
Sampled 1132 rows from: E:\Data\_Mining\_project\Data mining data\0724\_UkraineCombinedTweetsDeduped.csv  
Sampled 1208 rows from: E:\Data\_Mining\_project\Data mining data\0725\_UkraineCombinedTweetsDeduped.csv  
Sampled 1197 rows from: E:\Data\_Mining\_project\Data mining data\0726\_UkraineCombinedTweetsDeduped.csv  
Sampled 1478 rows from: E:\Data\_Mining\_project\Data mining data\0727\_UkraineCombinedTweetsDeduped.csv  
Sampled 1682 rows from: E:\Data\_Mining\_project\Data mining data\0728\_UkraineCombinedTweetsDeduped.csv  
Sampled 2215 rows from: E:\Data\_Mining\_project\Data mining data\0729\_UkraineCombinedTweetsDeduped.csv  
Sampled 1671 rows from: E:\Data\_Mining\_project\Data mining data\0730\_UkraineCombinedTweetsDeduped.csv  
Sampled 1862 rows from: E:\Data\_Mining\_project\Data mining data\0731\_UkraineCombinedTweetsDeduped.csv  
Sampled 1911 rows from: E:\Data\_Mining\_project\Data mining data\0801\_UkraineCombinedTweetsDeduped.csv  
Sampled 2859 rows from: E:\Data\_Mining\_project\Data mining data\0802\_UkraineCombinedTweetsDeduped.csv  
Sampled 2995 rows from: E:\Data\_Mining\_project\Data mining data\0803\_UkraineCombinedTweetsDeduped.csv  
Sampled 1791 rows from: E:\Data\_Mining\_project\Data mining data\0804\_UkraineCombinedTweetsDeduped.csv  
Sampled 1593 rows from: E:\Data\_Mining\_project\Data mining data\0805\_UkraineCombinedTweetsDeduped.csv  
Sampled 2047 rows from: E:\Data\_Mining\_project\Data mining data\0806\_UkraineCombinedTweetsDeduped.csv  
Sampled 2629 rows from: E:\Data\_Mining\_project\Data mining data\0807\_UkraineCombinedTweetsDeduped.csv  
Sampled 684 rows from: E:\Data\_Mining\_project\Data mining data\0808\_UkraineCombinedTweetsDeduped.csv  
Sampled 388 rows from: E:\Data\_Mining\_project\Data mining data\0809\_UkraineCombinedTweetsDeduped.csv  
Sampled 490 rows from: E:\Data\_Mining\_project\Data mining data\0810\_UkraineCombinedTweetsDeduped.csv  
Sampled 515 rows from: E:\Data\_Mining\_project\Data mining data\0811\_UkraineCombinedTweetsDeduped.csv

Sampled 597 rows from: E:\Data\_Mining\_project\Data mining data\0812\_UkraineCombinedTweetsDeduped.csv  
Sampled 410 rows from: E:\Data\_Mining\_project\Data mining data\0813\_UkraineCombinedTweetsDeduped.csv  
Sampled 332 rows from: E:\Data\_Mining\_project\Data mining data\0814\_UkraineCombinedTweetsDeduped.csv  
Sampled 391 rows from: E:\Data\_Mining\_project\Data mining data\0815\_UkraineCombinedTweetsDeduped.csv  
Sampled 448 rows from: E:\Data\_Mining\_project\Data mining data\0816\_UkraineCombinedTweetsDeduped.csv  
Sampled 556 rows from: E:\Data\_Mining\_project\Data mining data\0817\_UkraineCombinedTweetsDeduped.csv  
Sampled 481 rows from: E:\Data\_Mining\_project\Data mining data\0818\_UkraineCombinedTweetsDeduped.csv  
Sampled 480 rows from: E:\Data\_Mining\_project\Data mining data\0819\_UkraineCombinedTweetsDeduped.csv  
Sampled 444 rows from: E:\Data\_Mining\_project\Data mining data\0820\_UkraineCombinedTweetsDeduped.csv  
Sampled 477 rows from: E:\Data\_Mining\_project\Data mining data\0821\_UkraineCombinedTweetsDeduped.csv  
Sampled 470 rows from: E:\Data\_Mining\_project\Data mining data\0822\_UkraineCombinedTweetsDeduped.csv  
Sampled 503 rows from: E:\Data\_Mining\_project\Data mining data\0823\_UkraineCombinedTweetsDeduped.csv  
Sampled 797 rows from: E:\Data\_Mining\_project\Data mining data\0824\_UkraineCombinedTweetsDeduped.csv  
Sampled 585 rows from: E:\Data\_Mining\_project\Data mining data\0825\_UkraineCombinedTweetsDeduped.csv  
Sampled 555 rows from: E:\Data\_Mining\_project\Data mining data\0826\_UkraineCombinedTweetsDeduped.csv  
Sampled 370 rows from: E:\Data\_Mining\_project\Data mining data\0827\_UkraineCombinedTweetsDeduped.csv  
Sampled 370 rows from: E:\Data\_Mining\_project\Data mining data\0828\_UkraineCombinedTweetsDeduped.csv  
Sampled 534 rows from: E:\Data\_Mining\_project\Data mining data\0829\_UkraineCombinedTweetsDeduped.csv  
Sampled 631 rows from: E:\Data\_Mining\_project\Data mining data\0830\_UkraineCombinedTweetsDeduped.csv  
Sampled 538 rows from: E:\Data\_Mining\_project\Data mining data\0831\_UkraineCombinedTweetsDeduped.csv  
Sampled 483 rows from: E:\Data\_Mining\_project\Data mining data\0901\_UkraineCombinedTweetsDeduped.csv  
Sampled 537 rows from: E:\Data\_Mining\_project\Data mining data\0902\_UkraineCombinedTweetsDeduped.csv  
Sampled 407 rows from: E:\Data\_Mining\_project\Data mining data\0903\_UkraineCombinedTweetsDeduped.csv  
Sampled 402 rows from: E:\Data\_Mining\_project\Data mining data\0904\_UkraineCombinedTweetsDeduped.csv



Sampled 447 rows from: E:\Data\_Mining\_project\Data mining data\0905\_UkraineCombinedTweetsDeduped.csv

Sampled 616 rows from: E:\Data\_Mining\_project\Data mining data\0906\_UkraineCombinedTweetsDeduped.csv

Sampled 465 rows from: E:\Data\_Mining\_project\Data mining data\0907\_UkraineCombinedTweetsDeduped.csv

Sampled 478 rows from: E:\Data\_Mining\_project\Data mining data\0908\_UkraineCombinedTweetsDeduped.csv

Sampled 704 rows from: E:\Data\_Mining\_project\Data mining data\0909\_UkraineCombinedTweetsDeduped.csv

Sampled 650 rows from: E:\Data\_Mining\_project\Data mining data\0910\_UkraineCombinedTweetsDeduped.csv

Sampled 675 rows from: E:\Data\_Mining\_project\Data mining data\0911\_UkraineCombinedTweetsDeduped.csv

Sampled 629 rows from: E:\Data\_Mining\_project\Data mining data\0912\_UkraineCombinedTweetsDeduped.csv

Sampled 639 rows from: E:\Data\_Mining\_project\Data mining data\0913\_UkraineCombinedTweetsDeduped.csv

Sampled 754 rows from: E:\Data\_Mining\_project\Data mining data\0914\_UkraineCombinedTweetsDeduped.csv

Sampled 672 rows from: E:\Data\_Mining\_project\Data mining data\0915\_UkraineCombinedTweetsDeduped.csv

Sampled 631 rows from: E:\Data\_Mining\_project\Data mining data\0916\_UkraineCombinedTweetsDeduped.csv

Sampled 496 rows from: E:\Data\_Mining\_project\Data mining data\0917\_UkraineCombinedTweetsDeduped.csv

Sampled 424 rows from: E:\Data\_Mining\_project\Data mining data\0918\_UkraineCombinedTweetsDeduped.csv

Sampled 493 rows from: E:\Data\_Mining\_project\Data mining data\0919\_UkraineCombinedTweetsDeduped.csv

Sampled 574 rows from: E:\Data\_Mining\_project\Data mining data\0920\_UkraineCombinedTweetsDeduped.csv

Sampled 1043 rows from: E:\Data\_Mining\_project\Data mining data\0921\_UkraineCombinedTweetsDeduped.csv

Sampled 785 rows from: E:\Data\_Mining\_project\Data mining data\0922\_UkraineCombinedTweetsDeduped.csv

Sampled 774 rows from: E:\Data\_Mining\_project\Data mining data\0923\_UkraineCombinedTweetsDeduped.csv

Sampled 639 rows from: E:\Data\_Mining\_project\Data mining data\0924\_UkraineCombinedTweetsDeduped.csv

Sampled 530 rows from: E:\Data\_Mining\_project\Data mining data\0925\_UkraineCombinedTweetsDeduped.csv

Sampled 606 rows from: E:\Data\_Mining\_project\Data mining data\0926\_UkraineCombinedTweetsDeduped.csv

Sampled 689 rows from: E:\Data\_Mining\_project\Data mining data\0927\_UkraineCombinedTweetsDeduped.csv

Sampled 805 rows from: E:\Data\_Mining\_project\Data mining data\0928\_UkraineCombinedTweetsDeduped.csv

Sampled 670 rows from: E:\Data\_Mining\_project\Data mining data\0929\_UkraineCombinedTweetsDeduped.csv  
Sampled 699 rows from: E:\Data\_Mining\_project\Data mining data\0930\_UkraineCombinedTweetsDeduped.csv  
Sampled 657 rows from: E:\Data\_Mining\_project\Data mining data\1001\_UkraineCombinedTweetsDeduped.csv  
Sampled 512 rows from: E:\Data\_Mining\_project\Data mining data\1002\_UkraineCombinedTweetsDeduped.csv  
Sampled 600 rows from: E:\Data\_Mining\_project\Data mining data\1003\_UkraineCombinedTweetsDeduped.csv  
Sampled 593 rows from: E:\Data\_Mining\_project\Data mining data\1004\_UkraineCombinedTweetsDeduped.csv  
Sampled 539 rows from: E:\Data\_Mining\_project\Data mining data\1005\_UkraineCombinedTweetsDeduped.csv  
Sampled 554 rows from: E:\Data\_Mining\_project\Data mining data\1006\_UkraineCombinedTweetsDeduped.csv  
Sampled 601 rows from: E:\Data\_Mining\_project\Data mining data\1007\_UkraineCombinedTweetsDeduped.csv  
Sampled 865 rows from: E:\Data\_Mining\_project\Data mining data\1008\_UkraineCombinedTweetsDeduped.csv  
Sampled 664 rows from: E:\Data\_Mining\_project\Data mining data\1009\_UkraineCombinedTweetsDeduped.csv  
Sampled 1538 rows from: E:\Data\_Mining\_project\Data mining data\1010\_UkraineCombinedTweetsDeduped.csv  
Sampled 843 rows from: E:\Data\_Mining\_project\Data mining data\1011\_UkraineCombinedTweetsDeduped.csv  
Sampled 721 rows from: E:\Data\_Mining\_project\Data mining data\1012\_UkraineCombinedTweetsDeduped.csv  
Sampled 591 rows from: E:\Data\_Mining\_project\Data mining data\1013\_UkraineCombinedTweetsDeduped.csv  
Sampled 515 rows from: E:\Data\_Mining\_project\Data mining data\1014\_UkraineCombinedTweetsDeduped.csv  
Sampled 468 rows from: E:\Data\_Mining\_project\Data mining data\1015\_UkraineCombinedTweetsDeduped.csv  
Sampled 430 rows from: E:\Data\_Mining\_project\Data mining data\1016\_UkraineCombinedTweetsDeduped.csv  
Sampled 650 rows from: E:\Data\_Mining\_project\Data mining data\1017\_UkraineCombinedTweetsDeduped.csv  
Sampled 695 rows from: E:\Data\_Mining\_project\Data mining data\1018\_UkraineCombinedTweetsDeduped.csv  
Sampled 630 rows from: E:\Data\_Mining\_project\Data mining data\1019\_UkraineCombinedTweetsDeduped.csv  
Sampled 519 rows from: E:\Data\_Mining\_project\Data mining data\1020\_UkraineCombinedTweetsDeduped.csv  
Sampled 457 rows from: E:\Data\_Mining\_project\Data mining data\1021\_UkraineCombinedTweetsDeduped.csv  
Sampled 423 rows from: E:\Data\_Mining\_project\Data mining data\1022\_UkraineCombinedTweetsDeduped.csv

Sampled 409 rows from: E:\Data\_Mining\_project\Data mining data\1023\_UkraineCombinedTweetsDeduped.csv  
Sampled 438 rows from: E:\Data\_Mining\_project\Data mining data\1024\_UkraineCombinedTweetsDeduped.csv  
Sampled 459 rows from: E:\Data\_Mining\_project\Data mining data\1025\_UkraineCombinedTweetsDeduped.csv  
Sampled 459 rows from: E:\Data\_Mining\_project\Data mining data\1026\_UkraineCombinedTweetsDeduped.csv  
Sampled 438 rows from: E:\Data\_Mining\_project\Data mining data\1027\_UkraineCombinedTweetsDeduped.csv  
Sampled 738 rows from: E:\Data\_Mining\_project\Data mining data\1028\_UkraineCombinedTweetsDeduped.csv  
Sampled 912 rows from: E:\Data\_Mining\_project\Data mining data\1029\_UkraineCombinedTweetsDeduped.csv  
Sampled 362 rows from: E:\Data\_Mining\_project\Data mining data\1030\_UkraineCombinedTweetsDeduped.csv  
Sampled 431 rows from: E:\Data\_Mining\_project\Data mining data\1031\_UkraineCombinedTweetsDeduped.csv  
Sampled 461 rows from: E:\Data\_Mining\_project\Data mining data\1101\_UkraineCombinedTweetsDeduped.csv  
Sampled 463 rows from: E:\Data\_Mining\_project\Data mining data\1102\_UkraineCombinedTweetsDeduped.csv  
Sampled 459 rows from: E:\Data\_Mining\_project\Data mining data\1103\_UkraineCombinedTweetsDeduped.csv  
Sampled 483 rows from: E:\Data\_Mining\_project\Data mining data\1104\_UkraineCombinedTweetsDeduped.csv  
Sampled 451 rows from: E:\Data\_Mining\_project\Data mining data\1105\_UkraineCombinedTweetsDeduped.csv  
Sampled 415 rows from: E:\Data\_Mining\_project\Data mining data\1106\_UkraineCombinedTweetsDeduped.csv  
Sampled 614 rows from: E:\Data\_Mining\_project\Data mining data\1107\_UkraineCombinedTweetsDeduped.csv  
Sampled 438 rows from: E:\Data\_Mining\_project\Data mining data\1108\_UkraineCombinedTweetsDeduped.csv  
Sampled 636 rows from: E:\Data\_Mining\_project\Data mining data\1109\_UkraineCombinedTweetsDeduped.csv  
Sampled 509 rows from: E:\Data\_Mining\_project\Data mining data\1110\_UkraineCombinedTweetsDeduped.csv  
Sampled 609 rows from: E:\Data\_Mining\_project\Data mining data\1111\_UkraineCombinedTweetsDeduped.csv  
Sampled 476 rows from: E:\Data\_Mining\_project\Data mining data\1112\_UkraineCombinedTweetsDeduped.csv  
Sampled 427 rows from: E:\Data\_Mining\_project\Data mining data\1113\_UkraineCombinedTweetsDeduped.csv  
Sampled 490 rows from: E:\Data\_Mining\_project\Data mining data\1114\_UkraineCombinedTweetsDeduped.csv  
Sampled 1005 rows from: E:\Data\_Mining\_project\Data mining data\1115\_UkraineCombinedTweetsDeduped.csv

Sampled 918 rows from: E:\Data\_Mining\_project\Data mining data\1116\_UkraineCombinedTweetsDeduped.csv  
Sampled 511 rows from: E:\Data\_Mining\_project\Data mining data\1117\_UkraineCombinedTweetsDeduped.csv  
Sampled 386 rows from: E:\Data\_Mining\_project\Data mining data\1118\_UkraineCombinedTweetsDeduped.csv  
Sampled 474 rows from: E:\Data\_Mining\_project\Data mining data\1119\_UkraineCombinedTweetsDeduped.csv  
Sampled 425 rows from: E:\Data\_Mining\_project\Data mining data\1120\_UkraineCombinedTweetsDeduped.csv  
Sampled 895 rows from: E:\Data\_Mining\_project\Data mining data\1121\_UkraineCombinedTweetsDeduped.csv  
Sampled 493 rows from: E:\Data\_Mining\_project\Data mining data\1122\_UkraineCombinedTweetsDeduped.csv  
Sampled 920 rows from: E:\Data\_Mining\_project\Data mining data\1123\_UkraineCombinedTweetsDeduped.csv  
Sampled 741 rows from: E:\Data\_Mining\_project\Data mining data\1124\_UkraineCombinedTweetsDeduped.csv  
Sampled 474 rows from: E:\Data\_Mining\_project\Data mining data\1125\_UkraineCombinedTweetsDeduped.csv  
Sampled 388 rows from: E:\Data\_Mining\_project\Data mining data\1126\_UkraineCombinedTweetsDeduped.csv  
Sampled 535 rows from: E:\Data\_Mining\_project\Data mining data\1127\_UkraineCombinedTweetsDeduped.csv  
Sampled 647 rows from: E:\Data\_Mining\_project\Data mining data\1128\_UkraineCombinedTweetsDeduped.csv  
Sampled 459 rows from: E:\Data\_Mining\_project\Data mining data\1129\_UkraineCombinedTweetsDeduped.csv  
Sampled 378 rows from: E:\Data\_Mining\_project\Data mining data\1130\_UkraineCombinedTweetsDeduped.csv  
Sampled 611 rows from: E:\Data\_Mining\_project\Data mining data\1201\_UkraineCombinedTweetsDeduped.csv  
Sampled 709 rows from: E:\Data\_Mining\_project\Data mining data\1202\_UkraineCombinedTweetsDeduped.csv  
Sampled 397 rows from: E:\Data\_Mining\_project\Data mining data\1203\_UkraineCombinedTweetsDeduped.csv  
Sampled 408 rows from: E:\Data\_Mining\_project\Data mining data\1204\_UkraineCombinedTweetsDeduped.csv  
Sampled 476 rows from: E:\Data\_Mining\_project\Data mining data\1205\_UkraineCombinedTweetsDeduped.csv  
Sampled 473 rows from: E:\Data\_Mining\_project\Data mining data\1206\_UkraineCombinedTweetsDeduped.csv  
Sampled 472 rows from: E:\Data\_Mining\_project\Data mining data\1207\_UkraineCombinedTweetsDeduped.csv  
Sampled 596 rows from: E:\Data\_Mining\_project\Data mining data\1208\_UkraineCombinedTweetsDeduped.csv  
Sampled 429 rows from: E:\Data\_Mining\_project\Data mining data\1209\_UkraineCombinedTweetsDeduped.csv

Sampled 396 rows from: E:\Data\_Mining\_project\Data mining data\1210\_UkraineCombinedTweetsDeduped.csv  
Sampled 430 rows from: E:\Data\_Mining\_project\Data mining data\1211\_UkraineCombinedTweetsDeduped.csv  
Sampled 439 rows from: E:\Data\_Mining\_project\Data mining data\1212\_UkraineCombinedTweetsDeduped.csv  
Sampled 429 rows from: E:\Data\_Mining\_project\Data mining data\1213\_UkraineCombinedTweetsDeduped.csv  
Sampled 494 rows from: E:\Data\_Mining\_project\Data mining data\1214\_UkraineCombinedTweetsDeduped.csv  
Sampled 472 rows from: E:\Data\_Mining\_project\Data mining data\1215\_UkraineCombinedTweetsDeduped.csv  
Sampled 477 rows from: E:\Data\_Mining\_project\Data mining data\1216\_UkraineCombinedTweetsDeduped.csv  
Sampled 413 rows from: E:\Data\_Mining\_project\Data mining data\1217\_UkraineCombinedTweetsDeduped.csv  
Sampled 635 rows from: E:\Data\_Mining\_project\Data mining data\1218\_UkraineCombinedTweetsDeduped.csv  
Sampled 424 rows from: E:\Data\_Mining\_project\Data mining data\1219\_UkraineCombinedTweetsDeduped.csv  
Sampled 412 rows from: E:\Data\_Mining\_project\Data mining data\1220\_UkraineCombinedTweetsDeduped.csv  
Sampled 572 rows from: E:\Data\_Mining\_project\Data mining data\1221\_UkraineCombinedTweetsDeduped.csv  
Sampled 824 rows from: E:\Data\_Mining\_project\Data mining data\1222\_UkraineCombinedTweetsDeduped.csv  
Sampled 578 rows from: E:\Data\_Mining\_project\Data mining data\1223\_UkraineCombinedTweetsDeduped.csv  
Sampled 458 rows from: E:\Data\_Mining\_project\Data mining data\1224\_UkraineCombinedTweetsDeduped.csv  
Sampled 692 rows from: E:\Data\_Mining\_project\Data mining data\1225\_UkraineCombinedTweetsDeduped.csv  
Sampled 427 rows from: E:\Data\_Mining\_project\Data mining data\1226\_UkraineCombinedTweetsDeduped.csv  
Sampled 506 rows from: E:\Data\_Mining\_project\Data mining data\1227\_UkraineCombinedTweetsDeduped.csv  
Sampled 497 rows from: E:\Data\_Mining\_project\Data mining data\1228\_UkraineCombinedTweetsDeduped.csv  
Sampled 598 rows from: E:\Data\_Mining\_project\Data mining data\1229\_UkraineCombinedTweetsDeduped.csv  
Sampled 527 rows from: E:\Data\_Mining\_project\Data mining data\1230\_UkraineCombinedTweetsDeduped.csv  
Sampled 426 rows from: E:\Data\_Mining\_project\Data mining data\1231\_UkraineCombinedTweetsDeduped.csv  
Sampled 464 rows from: E:\Data\_Mining\_project\Data mining data\20230101\_UkraineCombinedTweetsDeduped.csv  
Sampled 464 rows from: E:\Data\_Mining\_project\Data mining data\20230102\_UkraineCombinedTweetsDeduped.csv

Sampled 395 rows from: E:\Data\_Mining\_project\Data mining data\20230103\_UkraineCombinedTweetsDeduped.csv  
Sampled 397 rows from: E:\Data\_Mining\_project\Data mining data\20230104\_UkraineCombinedTweetsDeduped.csv  
Sampled 452 rows from: E:\Data\_Mining\_project\Data mining data\20230105\_UkraineCombinedTweetsDeduped.csv  
Sampled 463 rows from: E:\Data\_Mining\_project\Data mining data\20230106\_UkraineCombinedTweetsDeduped.csv  
Sampled 374 rows from: E:\Data\_Mining\_project\Data mining data\20230107\_UkraineCombinedTweetsDeduped.csv  
Sampled 320 rows from: E:\Data\_Mining\_project\Data mining data\20230108\_UkraineCombinedTweetsDeduped.csv  
Sampled 394 rows from: E:\Data\_Mining\_project\Data mining data\20230109\_UkraineCombinedTweetsDeduped.csv  
Sampled 416 rows from: E:\Data\_Mining\_project\Data mining data\20230110\_UkraineCombinedTweetsDeduped.csv  
Sampled 429 rows from: E:\Data\_Mining\_project\Data mining data\20230111\_UkraineCombinedTweetsDeduped.csv  
Sampled 518 rows from: E:\Data\_Mining\_project\Data mining data\20230112\_UkraineCombinedTweetsDeduped.csv  
Sampled 550 rows from: E:\Data\_Mining\_project\Data mining data\20230113\_UkraineCombinedTweetsDeduped.csv  
Sampled 545 rows from: E:\Data\_Mining\_project\Data mining data\20230114\_UkraineCombinedTweetsDeduped.csv  
Sampled 664 rows from: E:\Data\_Mining\_project\Data mining data\20230115\_UkraineCombinedTweetsDeduped.csv  
Sampled 502 rows from: E:\Data\_Mining\_project\Data mining data\20230116\_UkraineCombinedTweetsDeduped.csv  
Sampled 428 rows from: E:\Data\_Mining\_project\Data mining data\20230117\_UkraineCombinedTweetsDeduped.csv  
Sampled 500 rows from: E:\Data\_Mining\_project\Data mining data\20230118\_UkraineCombinedTweetsDeduped.csv  
Sampled 441 rows from: E:\Data\_Mining\_project\Data mining data\20230119\_UkraineCombinedTweetsDeduped.csv  
Sampled 507 rows from: E:\Data\_Mining\_project\Data mining data\20230120\_UkraineCombinedTweetsDeduped.csv  
Sampled 470 rows from: E:\Data\_Mining\_project\Data mining data\20230121\_UkraineCombinedTweetsDeduped.csv  
Sampled 407 rows from: E:\Data\_Mining\_project\Data mining data\20230122\_UkraineCombinedTweetsDeduped.csv  
Sampled 430 rows from: E:\Data\_Mining\_project\Data mining data\20230123\_UkraineCombinedTweetsDeduped.csv  
Sampled 566 rows from: E:\Data\_Mining\_project\Data mining data\20230124\_UkraineCombinedTweetsDeduped.csv  
Sampled 697 rows from: E:\Data\_Mining\_project\Data mining data\20230125\_UkraineCombinedTweetsDeduped.csv  
Sampled 555 rows from: E:\Data\_Mining\_project\Data mining data\20230126\_UkraineCombinedTweetsDeduped.csv

Sampled 481 rows from: E:\Data\_Mining\_project\Data mining data\20230127\_UkraineCombinedTweetsDeduped.csv  
Sampled 395 rows from: E:\Data\_Mining\_project\Data mining data\20230128\_UkraineCombinedTweetsDeduped.csv  
Sampled 400 rows from: E:\Data\_Mining\_project\Data mining data\20230129\_UkraineCombinedTweetsDeduped.csv  
Sampled 417 rows from: E:\Data\_Mining\_project\Data mining data\20230130\_UkraineCombinedTweetsDeduped.csv  
Sampled 347 rows from: E:\Data\_Mining\_project\Data mining data\20230131\_UkraineCombinedTweetsDeduped.csv  
Sampled 433 rows from: E:\Data\_Mining\_project\Data mining data\20230201\_UkraineCombinedTweetsDeduped.csv  
Sampled 466 rows from: E:\Data\_Mining\_project\Data mining data\20230202\_UkraineCombinedTweetsDeduped.csv  
Sampled 425 rows from: E:\Data\_Mining\_project\Data mining data\20230203\_UkraineCombinedTweetsDeduped.csv  
Sampled 371 rows from: E:\Data\_Mining\_project\Data mining data\20230204\_UkraineCombinedTweetsDeduped.csv  
Sampled 389 rows from: E:\Data\_Mining\_project\Data mining data\20230205\_UkraineCombinedTweetsDeduped.csv  
Sampled 496 rows from: E:\Data\_Mining\_project\Data mining data\20230206\_UkraineCombinedTweetsDeduped.csv  
Sampled 816 rows from: E:\Data\_Mining\_project\Data mining data\20230207\_UkraineCombinedTweetsDeduped.csv  
Sampled 692 rows from: E:\Data\_Mining\_project\Data mining data\20230208\_UkraineCombinedTweetsDeduped.csv  
Sampled 968 rows from: E:\Data\_Mining\_project\Data mining data\20230209\_UkraineCombinedTweetsDeduped.csv  
Sampled 546 rows from: E:\Data\_Mining\_project\Data mining data\20230210\_UkraineCombinedTweetsDeduped.csv  
Sampled 475 rows from: E:\Data\_Mining\_project\Data mining data\20230211\_UkraineCombinedTweetsDeduped.csv  
Sampled 642 rows from: E:\Data\_Mining\_project\Data mining data\20230212\_UkraineCombinedTweetsDeduped.csv  
Sampled 485 rows from: E:\Data\_Mining\_project\Data mining data\20230213\_UkraineCombinedTweetsDeduped.csv  
Sampled 528 rows from: E:\Data\_Mining\_project\Data mining data\20230214\_UkraineCombinedTweetsDeduped.csv  
Sampled 515 rows from: E:\Data\_Mining\_project\Data mining data\20230215\_UkraineCombinedTweetsDeduped.csv  
Sampled 484 rows from: E:\Data\_Mining\_project\Data mining data\20230216\_UkraineCombinedTweetsDeduped.csv  
Sampled 543 rows from: E:\Data\_Mining\_project\Data mining data\20230217\_UkraineCombinedTweetsDeduped.csv  
Sampled 478 rows from: E:\Data\_Mining\_project\Data mining data\20230218\_UkraineCombinedTweetsDeduped.csv  
Sampled 507 rows from: E:\Data\_Mining\_project\Data mining data\20230219\_UkraineCombinedTweetsDeduped.csv

Sampled 838 rows from: E:\Data\_Mining\_project\Data mining data\20230220\_UkraineCombinedTweetsDeduped.csv

Sampled 1038 rows from: E:\Data\_Mining\_project\Data mining data\20230221\_UkraineCombinedTweetsDeduped.csv

Sampled 953 rows from: E:\Data\_Mining\_project\Data mining data\20230222\_UkraineCombinedTweetsDeduped.csv

Sampled 806 rows from: E:\Data\_Mining\_project\Data mining data\20230223\_UkraineCombinedTweetsDeduped.csv

Sampled 1256 rows from: E:\Data\_Mining\_project\Data mining data\20230224\_UkraineCombinedTweetsDeduped.csv

Sampled 2484 rows from: E:\Data\_Mining\_project\Data mining data\20230225\_UkraineCombinedTweetsDeduped.csv

Sampled 2134 rows from: E:\Data\_Mining\_project\Data mining data\20230226\_UkraineCombinedTweetsDeduped.csv

Sampled 2543 rows from: E:\Data\_Mining\_project\Data mining data\20230227\_UkraineCombinedTweetsDeduped.csv

Sampled 1334 rows from: E:\Data\_Mining\_project\Data mining data\20230228\_UkraineCombinedTweetsDeduped.csv

Sampled 1547 rows from: E:\Data\_Mining\_project\Data mining data\20230301\_UkraineCombinedTweetsDeduped.csv

Sampled 1632 rows from: E:\Data\_Mining\_project\Data mining data\20230302\_UkraineCombinedTweetsDeduped.csv

Sampled 1545 rows from: E:\Data\_Mining\_project\Data mining data\20230303\_UkraineCombinedTweetsDeduped.csv

Sampled 1379 rows from: E:\Data\_Mining\_project\Data mining data\20230304\_UkraineCombinedTweetsDeduped.csv

Sampled 1361 rows from: E:\Data\_Mining\_project\Data mining data\20230305\_UkraineCombinedTweetsDeduped.csv

Sampled 1601 rows from: E:\Data\_Mining\_project\Data mining data\20230306\_UkraineCombinedTweetsDeduped.csv

Sampled 1552 rows from: E:\Data\_Mining\_project\Data mining data\20230307\_UkraineCombinedTweetsDeduped.csv

Sampled 1704 rows from: E:\Data\_Mining\_project\Data mining data\20230308\_UkraineCombinedTweetsDeduped.csv

Sampled 1636 rows from: E:\Data\_Mining\_project\Data mining data\20230309\_UkraineCombinedTweetsDeduped.csv

Sampled 1203 rows from: E:\Data\_Mining\_project\Data mining data\20230310\_UkraineCombinedTweetsDeduped.csv

Sampled 1117 rows from: E:\Data\_Mining\_project\Data mining data\20230311\_UkraineCombinedTweetsDeduped.csv

Sampled 1207 rows from: E:\Data\_Mining\_project\Data mining data\20230312\_UkraineCombinedTweetsDeduped.csv

Sampled 1309 rows from: E:\Data\_Mining\_project\Data mining data\20230313\_UkraineCombinedTweetsDeduped.csv

Sampled 1352 rows from: E:\Data\_Mining\_project\Data mining data\20230314\_UkraineCombinedTweetsDeduped.csv

Sampled 1433 rows from: E:\Data\_Mining\_project\Data mining data\20230315\_UkraineCombinedTweetsDeduped.csv



Sampled 1438 rows from: E:\Data\_Mining\_project\Data mining data\20230316\_UkraineCombinedTweetsDeduped.csv  
Sampled 2083 rows from: E:\Data\_Mining\_project\Data mining data\20230317\_UkraineCombinedTweetsDeduped.csv  
Sampled 1843 rows from: E:\Data\_Mining\_project\Data mining data\20230318\_UkraineCombinedTweetsDeduped.csv  
Sampled 1759 rows from: E:\Data\_Mining\_project\Data mining data\20230319\_UkraineCombinedTweetsDeduped.csv  
Sampled 2022 rows from: E:\Data\_Mining\_project\Data mining data\20230320\_UkraineCombinedTweetsDeduped.csv  
Sampled 1795 rows from: E:\Data\_Mining\_project\Data mining data\20230321\_UkraineCombinedTweetsDeduped.csv  
Sampled 1840 rows from: E:\Data\_Mining\_project\Data mining data\20230322\_UkraineCombinedTweetsDeduped.csv  
Sampled 1586 rows from: E:\Data\_Mining\_project\Data mining data\20230323\_UkraineCombinedTweetsDeduped.csv  
Sampled 1556 rows from: E:\Data\_Mining\_project\Data mining data\20230324\_UkraineCombinedTweetsDeduped.csv  
Sampled 1362 rows from: E:\Data\_Mining\_project\Data mining data\20230325\_UkraineCombinedTweetsDeduped.csv  
Sampled 1331 rows from: E:\Data\_Mining\_project\Data mining data\20230326\_UkraineCombinedTweetsDeduped.csv  
Sampled 1308 rows from: E:\Data\_Mining\_project\Data mining data\20230327\_UkraineCombinedTweetsDeduped.csv  
Sampled 1254 rows from: E:\Data\_Mining\_project\Data mining data\20230328\_UkraineCombinedTweetsDeduped.csv  
Sampled 1366 rows from: E:\Data\_Mining\_project\Data mining data\20230329\_UkraineCombinedTweetsDeduped.csv  
Sampled 1816 rows from: E:\Data\_Mining\_project\Data mining data\20230330\_UkraineCombinedTweetsDeduped.csv  
Sampled 1205 rows from: E:\Data\_Mining\_project\Data mining data\20230331\_UkraineCombinedTweetsDeduped.csv  
Sampled 1335 rows from: E:\Data\_Mining\_project\Data mining data\20230401\_UkraineCombinedTweetsDeduped.csv  
Sampled 1358 rows from: E:\Data\_Mining\_project\Data mining data\20230402\_UkraineCombinedTweetsDeduped.csv  
Sampled 1204 rows from: E:\Data\_Mining\_project\Data mining data\20230403\_UkraineCombinedTweetsDeduped.csv  
Sampled 1495 rows from: E:\Data\_Mining\_project\Data mining data\20230404\_UkraineCombinedTweetsDeduped.csv  
Sampled 1564 rows from: E:\Data\_Mining\_project\Data mining data\20230405\_UkraineCombinedTweetsDeduped.csv  
Sampled 1363 rows from: E:\Data\_Mining\_project\Data mining data\20230406\_UkraineCombinedTweetsDeduped.csv  
Sampled 1222 rows from: E:\Data\_Mining\_project\Data mining data\20230407\_UkraineCombinedTweetsDeduped.csv  
Sampled 1062 rows from: E:\Data\_Mining\_project\Data mining data\20230408\_UkraineCombinedTweetsDeduped.csv

Sampled 1029 rows from: E:\Data\_Mining\_project\Data mining data\20230409\_UkraineCombinedTweetsDeduped.csv

Sampled 1203 rows from: E:\Data\_Mining\_project\Data mining data\20230410\_UkraineCombinedTweetsDeduped.csv

Sampled 1293 rows from: E:\Data\_Mining\_project\Data mining data\20230411\_UkraineCombinedTweetsDeduped.csv

Sampled 1403 rows from: E:\Data\_Mining\_project\Data mining data\20230412\_UkraineCombinedTweetsDeduped.csv

Sampled 1214 rows from: E:\Data\_Mining\_project\Data mining data\20230413\_UkraineCombinedTweetsDeduped.csv

Sampled 1131 rows from: E:\Data\_Mining\_project\Data mining data\20230414\_UkraineCombinedTweetsDeduped.csv

Sampled 1126 rows from: E:\Data\_Mining\_project\Data mining data\20230415\_UkraineCombinedTweetsDeduped.csv

Sampled 1078 rows from: E:\Data\_Mining\_project\Data mining data\20230416\_UkraineCombinedTweetsDeduped.csv

Sampled 1100 rows from: E:\Data\_Mining\_project\Data mining data\20230417\_UkraineCombinedTweetsDeduped.csv

Sampled 948 rows from: E:\Data\_Mining\_project\Data mining data\20230418\_UkraineCombinedTweetsDeduped.csv

Sampled 413 rows from: E:\Data\_Mining\_project\Data mining data\20230419\_UkraineCombinedTweetsDeduped.csv

Sampled 438 rows from: E:\Data\_Mining\_project\Data mining data\20230420\_UkraineCombinedTweetsDeduped.csv

Sampled 405 rows from: E:\Data\_Mining\_project\Data mining data\20230421\_UkraineCombinedTweetsDeduped.csv

Sampled 478 rows from: E:\Data\_Mining\_project\Data mining data\20230422\_UkraineCombinedTweetsDeduped.csv

Sampled 395 rows from: E:\Data\_Mining\_project\Data mining data\20230423\_UkraineCombinedTweetsDeduped.csv

Sampled 394 rows from: E:\Data\_Mining\_project\Data mining data\20230424\_UkraineCombinedTweetsDeduped.csv

Sampled 260 rows from: E:\Data\_Mining\_project\Data mining data\20230425\_UkraineCombinedTweetsDeduped.csv

Sampled 265 rows from: E:\Data\_Mining\_project\Data mining data\20230426\_UkraineCombinedTweetsDeduped.csv

Sampled 239 rows from: E:\Data\_Mining\_project\Data mining data\20230427\_UkraineCombinedTweetsDeduped.csv

Sampled 231 rows from: E:\Data\_Mining\_project\Data mining data\20230428\_UkraineCombinedTweetsDeduped.csv

Sampled 381 rows from: E:\Data\_Mining\_project\Data mining data\20230429\_to\_20230430\_UkraineCombinedTweetsDeduped.csv

Sampled 465 rows from: E:\Data\_Mining\_project\Data mining data\20230501\_to\_20230502\_UkraineCombinedTweetsDeduped.csv

Sampled 321 rows from: E:\Data\_Mining\_project\Data mining data\20230503\_UkraineCombinedTweetsDeduped.csv

Sampled 355 rows from: E:\Data\_Mining\_project\Data mining data\20230504\_UkraineCombinedTweetsDeduped.csv

Sampled 1244 rows from: E:\Data\_Mining\_project\Data mining data\20230505\_to\_20230508\_UkraineCombinedTweetsDeduped.csv

Sampled 336 rows from: E:\Data\_Mining\_project\Data mining data\20230509\_UkraineCombinedTweetsDeduped.csv

Sampled 312 rows from: E:\Data\_Mining\_project\Data mining data\20230510\_UkraineCombinedTweetsDeduped.csv

Sampled 604 rows from: E:\Data\_Mining\_project\Data mining data\20230511\_to\_20230512\_UkraineCombinedTweetsDeduped.csv

Sampled 509 rows from: E:\Data\_Mining\_project\Data mining data\20230513\_UkraineCombinedTweetsDeduped.csv

Sampled 469 rows from: E:\Data\_Mining\_project\Data mining data\20230514\_UkraineCombinedTweetsDeduped.csv

Sampled 691 rows from: E:\Data\_Mining\_project\Data mining data\20230515\_to\_20230516\_UkraineCombinedTweetsDeduped.csv

Sampled 256 rows from: E:\Data\_Mining\_project\Data mining data\20230517\_UkraineCombinedTweetsDeduped.csv

Sampled 1326 rows from: E:\Data\_Mining\_project\Data mining data\20230518\_to\_20230520\_UkraineCombinedTweetsDeduped.csv

Sampled 521 rows from: E:\Data\_Mining\_project\Data mining data\20230521\_UkraineCombinedTweetsDeduped.csv

Sampled 392 rows from: E:\Data\_Mining\_project\Data mining data\20230522\_UkraineCombinedTweetsDeduped.csv

Sampled 261 rows from: E:\Data\_Mining\_project\Data mining data\20230523\_UkraineCombinedTweetsDeduped.csv

Sampled 220 rows from: E:\Data\_Mining\_project\Data mining data\20230524\_UkraineCombinedTweetsDeduped.csv

Sampled 211 rows from: E:\Data\_Mining\_project\Data mining data\20230525\_UkraineCombinedTweetsDeduped.csv

Sampled 205 rows from: E:\Data\_Mining\_project\Data mining data\20230526\_UkraineCombinedTweetsDeduped.csv

Sampled 69 rows from: E:\Data\_Mining\_project\Data mining data\20230527\_UkraineCombinedTweetsDeduped.csv

Sampled 153 rows from: E:\Data\_Mining\_project\Data mining data\20230528\_UkraineCombinedTweetsDeduped.csv

Sampled 179 rows from: E:\Data\_Mining\_project\Data mining data\20230529\_UkraineCombinedTweetsDeduped.csv

Sampled 205 rows from: E:\Data\_Mining\_project\Data mining data\20230530\_UkraineCombinedTweetsDeduped.csv

Sampled 101 rows from: E:\Data\_Mining\_project\Data mining data\20230531\_UkraineCombinedTweetsDeduped.csv

Sampled 145 rows from: E:\Data\_Mining\_project\Data mining data\20230601\_UkraineCombinedTweetsDeduped.csv

Sampled 83 rows from: E:\Data\_Mining\_project\Data mining data\20230602\_UkraineCombinedTweetsDeduped.csv

Sampled 598 rows from: E:\Data\_Mining\_project\Data mining data\20230603\_UkraineCombinedTweetsDeduped.csv

Sampled 648 rows from: E:\Data\_Mining\_project\Data mining data\20230604\_UkraineCombinedTweetsDeduped.csv

Sampled 699 rows from: E:\Data\_Mining\_project\Data mining data\20230605\_UkraineCombinedTweetsDeduped.csv

Sampled 1183 rows from: E:\Data\_Mining\_project\Data mining data\20230606\_UkraineCombinedTweetsDeduped.csv

Sampled 250 rows from: E:\Data\_Mining\_project\Data mining data\20230607\_UkraineCombinedTweetsDeduped.csv

Sampled 1246 rows from: E:\Data\_Mining\_project\Data mining data\20230608\_UkraineCombinedTweetsDeduped.csv

Sampled 1166 rows from: E:\Data\_Mining\_project\Data mining data\20230609\_UkraineCombinedTweetsDeduped.csv

Sampled 854 rows from: E:\Data\_Mining\_project\Data mining data\20230610\_UkraineCombinedTweetsDeduped.csv

Sampled 873 rows from: E:\Data\_Mining\_project\Data mining data\20230611\_UkraineCombinedTweetsDeduped.csv

Sampled 1064 rows from: E:\Data\_Mining\_project\Data mining data\20230612\_UkraineCombinedTweetsDeduped.csv

Sampled 1240 rows from: E:\Data\_Mining\_project\Data mining data\20230613\_UkraineCombinedTweetsDeduped.csv

Sampled 762 rows from: E:\Data\_Mining\_project\Data mining data\20230614\_UkraineCombinedTweetsDeduped.csv

Sampled 12293 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped20220227-131611.csv

Sampled 3570 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_FEB27.csv

Sampled 2378 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_FEB28\_part1.csv

Sampled 1404 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_FEB28\_part2.csv

Sampled 4093 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR01.csv

Sampled 4170 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR02.csv

Sampled 3958 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR03.csv

Sampled 4803 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR04.csv

Sampled 5468 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR05.csv

Sampled 5668 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR06.csv

Sampled 5677 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR07.csv

Sampled 5194 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR08.csv

Sampled 4939 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR09.csv

Sampled 4601 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR10.csv

Sampled 4104 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR11.csv  
 Sampled 3979 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR12.csv  
 Sampled 4520 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR13.csv  
 Sampled 4285 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR14.csv  
 Sampled 4842 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR15.csv  
 Sampled 4409 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR16.csv  
 Sampled 4686 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR17.csv  
 Sampled 4681 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR18.csv  
 Sampled 4427 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR19.csv  
 Sampled 4233 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR20.csv  
 Sampled 4805 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR21.csv  
 Sampled 4131 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR22.csv  
 Sampled 4125 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR23.csv  
 Sampled 3778 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR24.csv  
 Sampled 4241 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR25.csv  
 Sampled 4113 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR26.csv  
 Sampled 7682 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR27\_to\_28.csv  
 Sampled 3641 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR29.csv  
 Sampled 3489 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR30.csv  
 Sampled 3445 rows from: E:\Data\_Mining\_project\Data mining data\UkraineCombinedTweetsDeduped\_MAR31.csv

Final sampled DataFrame shape: (708764, 28)

```
[3]: # Save to CSV
df.to_csv("E:/Data_Mining_project/sampled_dataset_1%.csv", index=False)
print(" Sampled dataset saved to: E:/Data_Mining_project/sampled_dataset.csv")
```

Sampled dataset saved to: E:/Data\_Mining\_project/sampled\_dataset.csv

```
[4]: # Print the number of rows and columns in the DataFrame  
print(df.shape)
```

```
(708764, 28)
```

```
[ ]:
```