

```
!pip install -q vaderSentiment transformers torch
```

```
import re
import pandas as pd
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from transformers import pipeline
```

_____	126.0/126.0	kB	3.1 MB/s	eta 0:00:00
_____	363.4/363.4	MB	3.5 MB/s	eta 0:00:00
_____	13.8/13.8	MB	40.9 MB/s	eta 0:00:00
_____	24.6/24.6	MB	48.2 MB/s	eta 0:00:00
_____	883.7/883.7	kB	23.7 MB/s	eta 0:00:00
_____	664.8/664.8	MB	2.6 MB/s	eta 0:00:00
_____	211.5/211.5	MB	5.7 MB/s	eta 0:00:00
_____	56.3/56.3	MB	16.2 MB/s	eta 0:00:00
_____	127.9/127.9	MB	8.8 MB/s	eta 0:00:00
_____	207.5/207.5	MB	5.7 MB/s	eta 0:00:00
_____	21.1/21.1	MB	37.7 MB/s	eta 0:00:00

```
from google.colab import files
uploaded = files.upload()

df = pd.read_csv('/content/tweets-data.csv')
df = df.sample(500, random_state=0).reset_index(drop=True)

print(df.columns.tolist())
df.head()
```

Sélect. fichiers

Aucun fichier choisi

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving tweets-data.csv to tweets-data (1).csv

['Unnamed: 0', 'Date Created', 'Number of Likes', 'Source of Tweet', 'Tweets', 'hashtag']

Unnamed: 0	Date Created	Number of Likes	Source of Tweet	Tweets	hashtag
0	830 2023-06-25 12:51:58+00:00	0	NaN	I must be old because I can't believe that the...	titanic
1	361 2023-06-25 18:43:48+00:00	0	NaN	Glastonbury 2023, Sunday live: Lil Nas X bring...	wagner
2	190 2023-06-25 17:57:04+00:00	4	NaN	Subscription Available👉 1 Day Free Trial👉 1...	titanic
3	243 2023-06-25 17:09:03+00:00	0	NaN	Hello everyone 🌎We are providing best iptv se...	titan

```
def clean_tweet(text: str) -> str:
    text = text.lower()
    text = re.sub(r'http\S+', '', text)
    text = re.sub(r'@\w+', '', text)
    text = re.sub(r'[^a-z\s]', '', text)
    tokens = text.split()
    return " ".join(tokens)

df['clean_text'] = df['Tweets'].map(clean_tweet)
df[['Tweets', 'clean_text']].head()
```

	Tweets	clean_text
0	I must be old because I can't believe that the...	i must be old because i cant believe that ther...
1	Glastonbury 2023, Sunday live: Lil Nas X bring...	glastonbury sunday live lil nas x brings the d...
2	Subscription Available👉 1 Day Free Trial👉 1...	subscription available day free trial klive ch...
3	Hello everyone 🌎We are providing best iptv se...	hello everyone we are providing best iptv serv...

```
# Initialise VADER
analyzer = SentimentIntensityAnalyzer()

def analyze_vader(text: str):
    comp = analyzer.polarity_scores(text)['compound']
```

```
label = 'pos' if comp>0 else 'neg' if comp<0 else 'neu'
return label, comp

df['vader_label'], df['vader_score'] = zip(*df['clean_text'].map(analyze_vader))

# Aperçu
df[['clean_text', 'vader_label', 'vader_score']].head()
```

↗

	clean_text	vader_label	vader_score
0	i must be old because i cant believe that ther...	neg	-0.5267
1	glastonbury sunday live lil nas x brings the d...	pos	0.3400
2	subscription available day free trial klive ch...	pos	0.6808
3	hello everyone we are providing best iptv serv...	pos	0.8555
4	hmmm there are people who may beg to differ titan	neu	0.0000

```
hf_pipe = pipeline("sentiment-analysis")
```

```
results = df['clean_text'].apply(lambda t: hf_pipe(t)[0])
df['hf_label'] = results.map(lambda r: r['label'])
df['hf_score'] = results.map(lambda r: r['score'])
```

```
# Aperçu
df[['clean_text', 'hf_label', 'hf_score']].head()
```

↗

No model was supplied, defaulted to distilbert/distilbert-base-uncased-finetuned-sst-2-english and revision 714eb0f (<https://huggingf>

Using a pipeline without specifying a model name and revision in production is not recommended.

/usr/local/lib/python3.11/dist-packages/huggingface\_hub/utils/\_auth.py:94: UserWarning:

The secret `HF\_TOKEN` does not exist in your Colab secrets.

To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), set it as se

You will be able to reuse this secret in all of your notebooks.

Please note that authentication is recommended but still optional to access public models or datasets.

warnings.warn(

config.json: 100% 629/629 [00:00<00:00, 17.8kB/s]

model.safetensors: 100% 268M/268M [00:03<00:00, 99.7MB/s]

tokenizer\_config.json: 100% 48.0/48.0 [00:00<00:00, 1.62kB/s]

vocab.txt: 232k/? [00:00<00:00, 5.81MB/s]

Device set to use cpu

	clean_text	hf_label	hf_score
0	i must be old because i cant believe that ther...	NEGATIVE	0.993338
1	glastonbury sunday live lil nas x brings the d...	POSITIVE	0.999661
2	subscription available day free trial klive ch...	NEGATIVE	0.984930
3	hello everyone we are providing best iptv serv...	POSITIVE	0.569796
4	hmmm there are people who may beg to differ titan	NEGATIVE	0.994560

```
print("Répartition VADER :")
print(df['vader_label'].value_counts(), "\n")

print("Répartition Transformers :")
print(df['hf_label'].value_counts())
```

↗

```
Répartition VADER :
vader_label
neu      255
pos      124
neg      121
Name: count, dtype: int64

Répartition Transformers :
hf_label
NEGATIVE    391
POSITIVE    109
Name: count, dtype: int64
```

```
df[['Tweets',
     'clean_text',
     'vader_label', 'vader_score',
     'hf_label', 'hf_score'
```

]].head(20)



	Tweets	clean_text	vader_label	vader_score	hf_label	hf_score
0	I must be old because I can't believe that the...	i must be old because i cant believe that ther...	neg	-0.5267	NEGATIVE	0.993338
1	Glastonbury 2023, Sunday live: Lil Nas X bring...	glastonbury sunday live lil nas x brings the d...	pos	0.3400	POSITIVE	0.999661
2	Subscription Available\n📺 1 Day Free Trial\n📺 1...	subscription available day free trial klive ch...	pos	0.6808	NEGATIVE	0.984930
3	Hello everyone \nWe are providing best iptv se...	hello everyone we are providing best iptv serv...	pos	0.8555	POSITIVE	0.569796
4	Hmmm, there are 5 people who may beg to differ...	hmmm there are people who may beg to differ titan	neu	0.0000	NEGATIVE	0.994560
5	🇪🇸 #Titán II La empresa #OceanGate emitió un com...	titn la empresa oceangate emiti un comunicado ...	neu	0.0000	NEGATIVE	0.991795
6	Putin: Rusia cumplirán todos los objetivos en ...	putin rusia cumplirn todos los objetivos en uc...	neu	0.0000	NEGATIVE	0.973084
7	ビットコイン一撃7300万円 <u>100</u> <u>100</u> <u>100</u> \n\nhttps://t.co/4nTlpt6ucK\n...	btc bitcoin fxgt glastonbury lgbt russia russi...	neu	0.0000	NEGATIVE	0.981829
8	@DrAmirKhan22 #Titanic #Titan #Messi🏆🏆🏆 #sundayv...	titanic titan messi sundayvibes	neu	0.0000	NEGATIVE	0.988846
9	#Wagner #Russia #Prighozin #Rostow #PatrickLan...	wagner russia prighozin rostow patricklancaster	neu	0.0000	POSITIVE	0.791998
10	Here is a picture of #Titan (previously named ...	here is a picture of titan previously named cy...	pos	0.3612	NEGATIVE	0.996040
11	Learnt my lesson thank you #Titan #Titanic #Su...	learnt my lesson thank you titan titanic subma...	pos	0.3612	POSITIVE	0.994111
12	タイタニック前編\n#Titanic \n#LeonardoDiCaprio\n#タイタニッ...	titanic leonardodicaprio	neu	0.0000	NEGATIVE	0.636090
13	#Titanic ... 4 dias. Quatro...	titanic dias quatro	neu	0.0000	NEGATIVE	0.976375
14	Abominable showman: the rise of #Wagner's Yevg...	abominable showman the rise of wagners yevgeny...	neu	0.0000	NEGATIVE	0.905801