📘 Sentiment Analysis on Tweets Using VADER and Transformers

🔍 Objective

The goal of this project is to perform **sentiment analysis** on a sample of tweets using:

- A **lexicon-based approach**: VADER (Valence Aware Dictionary and sEntiment Reasoner)

- A **model-based approach**: Transformer-based pre-trained model from Hugging Face

Both approaches output two elements:

- **Sentiment label** (positive / neutral / negative)

- **Sentiment score** (numeric)

📂 Dataset

- **Source**: `tweets-data.csv`

- **Sample Size**: 500 randomly selected tweets

- **Relevant Column**: `Tweets`

🛠️ Methodology

1. Data Cleaning

Each tweet was cleaned using the following steps:

- Converted to lowercase

- Removed URLs

- Removed special characters and numbers

- Removed stopwords

- Reconstructed the cleaned sentence

✅ Output: New column `cleaned_text`

2. VADER Sentiment Analysis

- Utilized `nltk.sentiment.SentimentIntensityAnalyzer`

- The compound score was used to assign labels:

  - `compound >= 0.05` → Positive

  - `compound <= -0.05` → Negative

  - Else → Neutral

✅ Output:

- `vader_label`

- `vader_score`

3. Transformers Sentiment Analysis

- Model: `distilbert-base-uncased-finetuned-sst-2-english` (via Hugging Face `pipeline`)

- Tweets longer than 512 characters were **truncated** before inference to prevent indexing errors.

- Output includes:

  - `transformer_label`: predicted label

- `transformer_score`: model confidence

✅ Output:

- `transformer_label`

- `transformer_score`

🧾 Sample Output

Tweet ID	Tweets	cleaned_text	vader_label	vader_score	transformer_label	transformer_score
2899	Le #DessinDePresse de Sanaga : ls sont morts comme des héros...	le dessindepresse de sanaga ls sont morts comme des héros...	positive	0.4767	negative	0.9815
594	#Russia #Wagner #RussiaCivilWar https://t.co/...	russia wagner russiacivil war	neutral	0.0000	negative	0.9620
2870	Exclusive content -https://t.co/...	exclusive content cosplay japan titan titanics...	negative	-0.4404	negative	0.9615

52	Auch heute geht die politische Nachricht des Tages...	auch heute geht die politische nachricht des tages...	negative	-0.5994	negative	0.9755
1391	@crazyclipsonly Same type that would take a homemade...	crazyclipsonly type would take homemade playstation ...	neutral	0.0000	negative	0.9916