```
Start coding or generate with AI.
```

Step 1:install and import libraries

```
!pip install nltk
import nltk
from nltk.tokenize import TweetTokenizer
from nltk.corpus import twitter_samples
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.12/dist-packages (from nltk) (8.3.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.12/dist-packages (from nltk) (1.5.3)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/dist-packages (from nltk) (2025.11.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from nltk) (4.67.1)
```

step 2:download required nltk resources

```
nltk.download('twitter_samples')
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger_eng')
```

```
[nltk_data] Downloading package twitter_samples to /root/nltk_data...
[nltk_data]   Package twitter_samples is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger_eng is already up-to-
[nltk_data]       date!
True
```

step 3:load tweets dataset

```
tweets=twitter_samples.strings('positive_tweets.json')
for i in range(3):
  print("Tweet",i+1)
  print(tweets[i])
  print()
```

```
Tweet 1
#FollowFriday @France_Inte @PKuchly57 @Milipol_Paris for being top engaged members in my community this week :)

Tweet 2
@Lamb2ja Hey James! How odd :/ Please call our Contact Centre on 02392441234 and we will be able to assist you :) Many thank

Tweet 3
@DespiteOfficial we had a listen last night :) As You Bleed is an amazing track. When are you in Scotland?!
```

step 4:tokenization using tweettokenizer

```
tokenizer=TweetTokenizer(
    preserve_case=False,
    strip_handles=True,
    reduce_len=True
)
tokenized_tweets=[tokenizer.tokenize(tweet) for tweet in tweets[:5]]
for i,tokens in enumerate(tokenized_tweets):
  print("tweet",i+1,"Tokens:")
  print(tokens)
  print()
```

```
tweet 1 Tokens:
['#followfriday', 'for', 'being', 'top', 'engaged', 'members', 'in', 'my', 'community', 'this', 'week', ':)']

tweet 2 Tokens:
['hey', 'james', '!', 'how', 'odd', ':/', 'please', 'call', 'our', 'contact', 'centre', 'on', '02392441234', 'and', 'we', 'w

tweet 3 Tokens:
['we', 'had', 'a', 'listen', 'last', 'night', ':)', 'as', 'you', 'bleed', 'is', 'an', 'amazing', 'track', '.', 'when', 'are'

tweet 4 Tokens:
['congrats', ':)']

tweet 5 Tokens:
```

```
['yeaaah', 'yipppy', '!', '!', '!', 'my', 'accnt', 'verified', 'rqst', 'has', 'succeed', 'got', 'a', 'blue', 'tick', 'mark',
```

```
Start coding or generate with AI.
```

step 3:pos tagging on custom noisy text

```
text="omg I luv this phone 😍 #awesome #AI"
tokens=tokenizer.tokenize(text)
tags=nltk.pos_tag(tokens)
print("orginal text:",text)
print("Tokens:",tokens)
print("POS Tags:",tags)
```

```
orginal text: omg I luv this phone 😍 #awesome #AI
Tokens: ['omg', 'i', 'luv', 'this', 'phone', '😍', '#awesome', '#ai']
POS Tags: [('omg', 'NN'), ('i', 'NN'), ('luv', 'VBP'), ('this', 'DT'), ('phone', 'NN'), ('😍', 'VBD'), ('#awesome', 'NNP'),
```

step 5:pos tagging using nltk

```
pos_tagged_tweets=[nltk.pos_tag(tokens) for tokens in tokenized_tweets]
for i,tagged in enumerate(pos_tagged_tweets):
  print("Tweet",i+1,"POS Tags:")
  print(tagged)
  print()
```

```
Tweet 1 POS Tags:
[('#followfriday', 'NN'), ('for', 'IN'), ('being', 'VBG'), ('top', 'JJ'), ('engaged', 'VBN'), ('members', 'NNS'), ('in', 'IN

Tweet 2 POS Tags:
[('hey', 'NN'), ('james', 'NNS'), ('!', '.'), ('how', 'WRB'), ('odd', 'JJ'), (':/', 'JJ'), ('please', 'NN'), ('call', 'VB'),

Tweet 3 POS Tags:
[('we', 'PRP'), ('had', 'VBD'), ('a', 'DT'), ('listen', 'VBN'), ('last', 'JJ'), ('night', 'NN'), (':)', 'NN'), ('as', 'IN'),

Tweet 4 POS Tags:
[('congrats', 'NNS'), (':)', 'VBP')]

Tweet 5 POS Tags:
[('yeaaah', 'NN'), ('yipppy', 'JJ'), ('!', '.'), ('!', '.'), ('!', '.'), ('my', 'PRP$'), ('accnt', 'JJ'), ('verified', 'VBN'
```

extract nouns and verbs

```
nouns=[]
verbs=[]
for word,tag in tags:
  if tag.startswith('NN'):
    nouns.append(word)
  elif tag.startswith('VB'):
    verbs.append(word)
print("nouns:",nouns)
print("verbs:",verbs)
```

```
nouns: ['omg', 'i', 'phone', '#awesome', '#ai']
verbs: ['luv', '😍']
```