```
import pandas as pd
import re


# Load the dataset
df = pd.read_csv("tweeter_dataset.csv", encoding="latin-1", header=None, names=["target", "ids", "date", "flag", "user", "text"])


df.head()
```

	target	ids	date	flag	user	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....

+ Code + Text

```
# Remove unnecessary columns
df = df.drop(["target", "ids", "date", "flag"], axis=1)


df.columns
```

```
Index(['user', 'text'], dtype='object')
```

```
duplicate_count = df.duplicated(subset=['text']).sum()

print("Number of Duplicate Records:", duplicate_count)
print(df.shape)
```

```
Number of Duplicate Records: 18534
(1600000, 2)
```

```
# Remove duplicate records
df = df.drop_duplicates(subset=['text'])

print(df.shape)
```

```
(1581466, 2)
```

```
#Check if dataset have missing values
df.isnull().sum()
```

```
user    0
text    0
dtype: int64
```

```
# Text cleaning
def clean_text(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text)  # Remove mentions
    text = re.sub('https?://[A-Za-z0-9./]+', '', text)  # Remove URLs
    text = re.sub("[^a-zA-Z]", " ", text)  # Remove special characters and numbers
    text = text.lower()  # Convert to lowercase
    return text

df['clean_text'] = df['text'].apply(clean_text)


df.head()
```

	user	text	clean_text
0	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...	awww that s a bummer you shoulda got da...
1	scotthamilton	is upset that he can't update his Facebook by ...	is upset that he can t update his facebook by ...
2	mattycus	@Kenichan I dived many times for the ball. Man...	i dived many times for the ball managed to s...
3	ElleCTF	my whole body feels itchy and like its on fire	my whole body feels itchy and like its on fire
4	Karoli	@nationwideclass no, it's not behaving at all....	no it s not behaving at all i m mad why am...

```
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer


nltk.download('vader_lexicon')
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\satre\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
True
```

```
# Check unique values in the 'target' column
print(df['target'].unique())



sia = SentimentIntensityAnalyzer()

# Apply sentiment analysis to the clean_text column
df['sentiment_score'] = df['clean_text'].apply(lambda x: sia.polarity_scores(x)['compound'])

# Convert the sentiment scores to categories (positive, negative, neutral)
df['sentiment'] = df['sentiment_score'].apply(lambda x: 'positive' if x > 0 else 'negative' if x < 0 else 'neutral')



df.head()
```

	user	text	clean_text	sentiment_score	sentiment
0	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...	awww that s a bummer you shoulda got da...	-0.3818	negative
1	scotthamilton	is upset that he can't update his Facebook by ...	is upset that he can t update his facebook by ...	-0.7269	negative
2	mattycus	@Kenichan I dived many times for the ball. Man...	i dived many times for the ball managed to s...	0.4939	positive
3	ElleCTF	my whole body feels itchy and like its on fire	my whole body feels itchy and like its on fire	-0.2500	negative
4	Karoli	@nationwideclass no, it's not behaving at all....	no it s not behaving at all i m mad why am...	-0.6597	negative