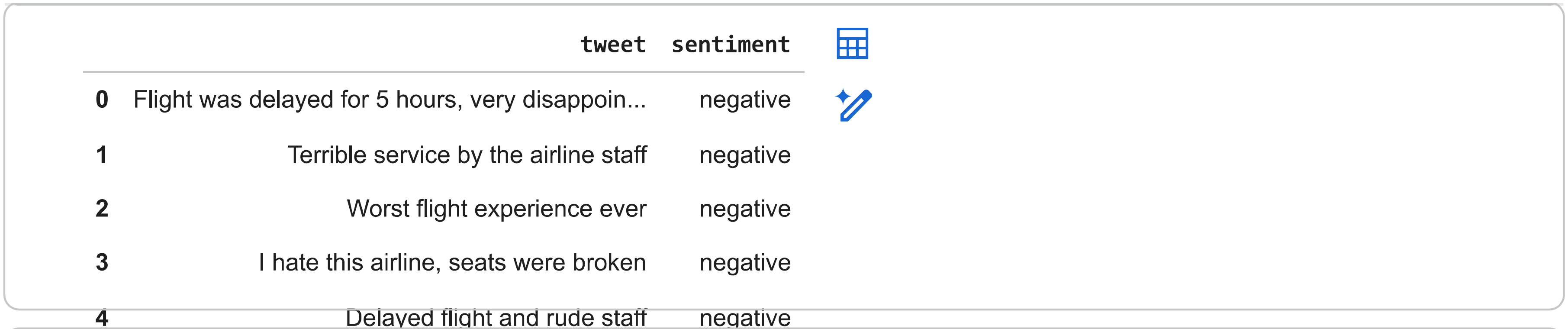
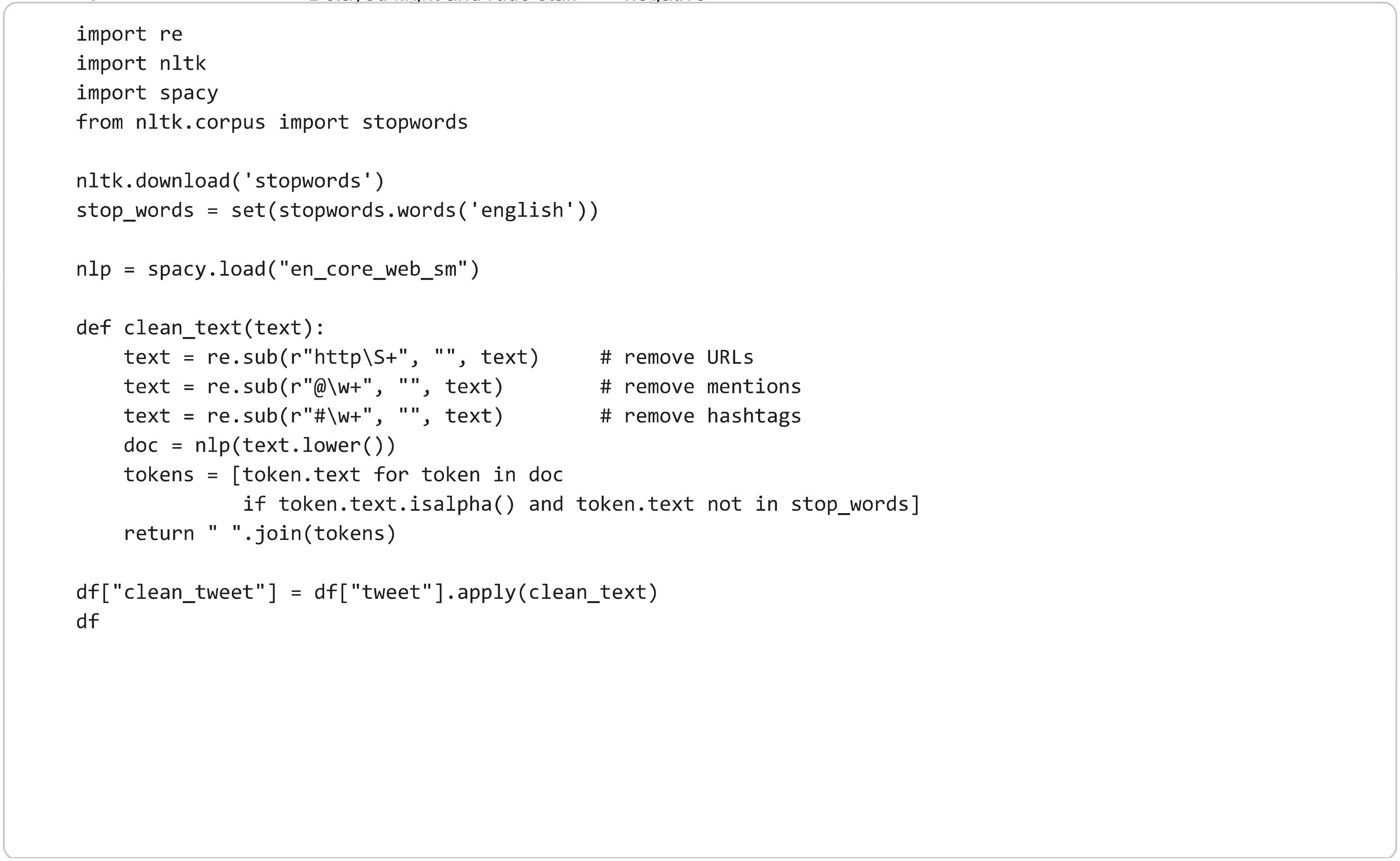
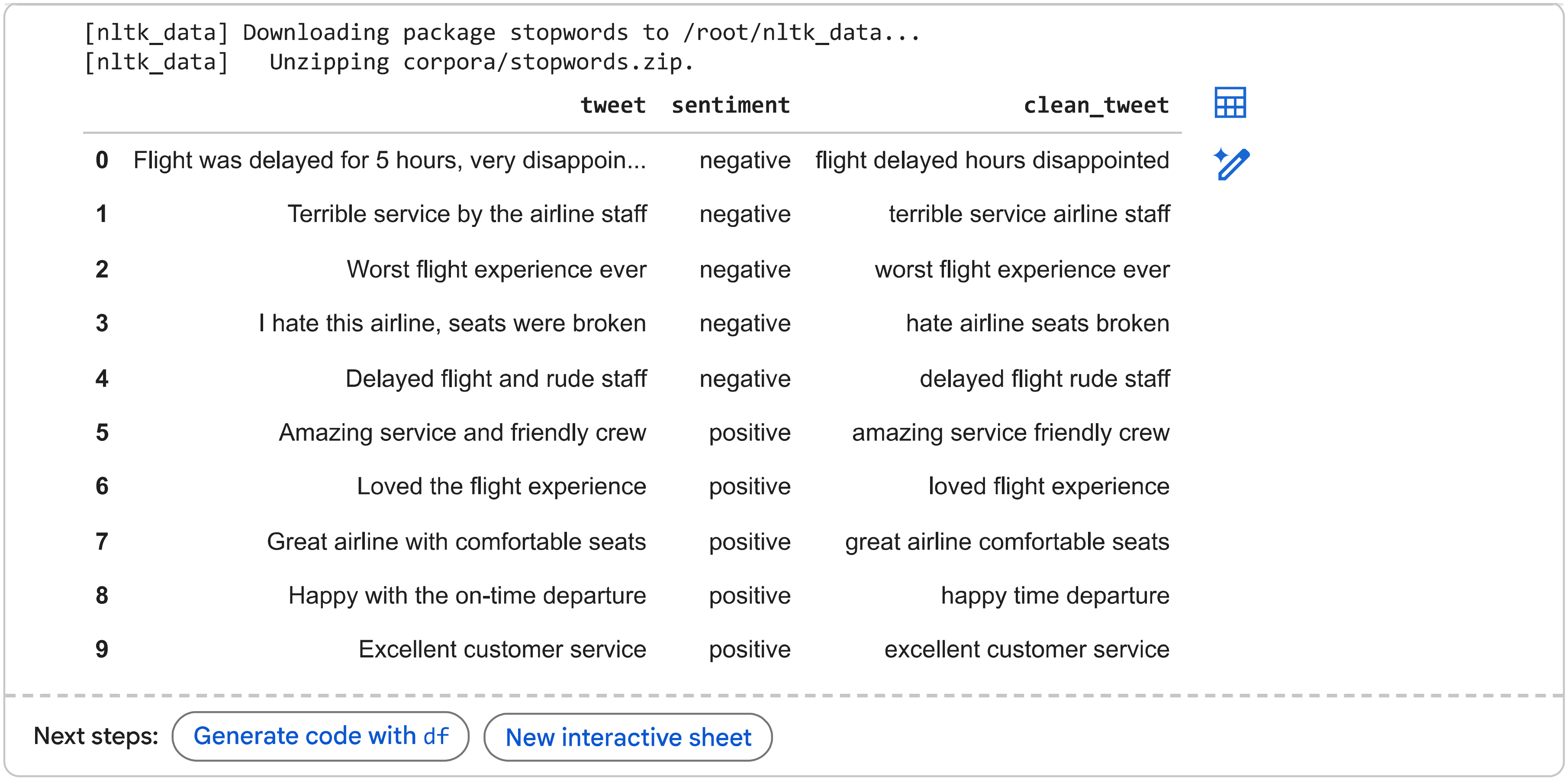
```python
import pandas as pd

data = {
    "tweet": [
        "Flight was delayed for 5 hours, very disappointed!",
        "Terrible service by the airline staff",
        "Worst flight experience ever",
        "I hate this airline, seats were broken",
        "Delayed flight and rude staff",
        "Amazing service and friendly crew",
        "Loved the flight experience",
        "Great airline with comfortable seats",
        "Happy with the on-time departure",
        "Excellent customer service"
    ],
    "sentiment": [
        "negative", "negative", "negative", "negative", "negative",
        "positive", "positive", "positive", "positive", "positive"
    ]
}

df = pd.DataFrame(data)
df
```



	tweet	sentiment
0	Flight was delayed for 5 hours, very disappoin...	negative
1	Terrible service by the airline staff	negative
2	Worst flight experience ever	negative
3	I hate this airline, seats were broken	negative
4	Delayed flight and rude staff	negative

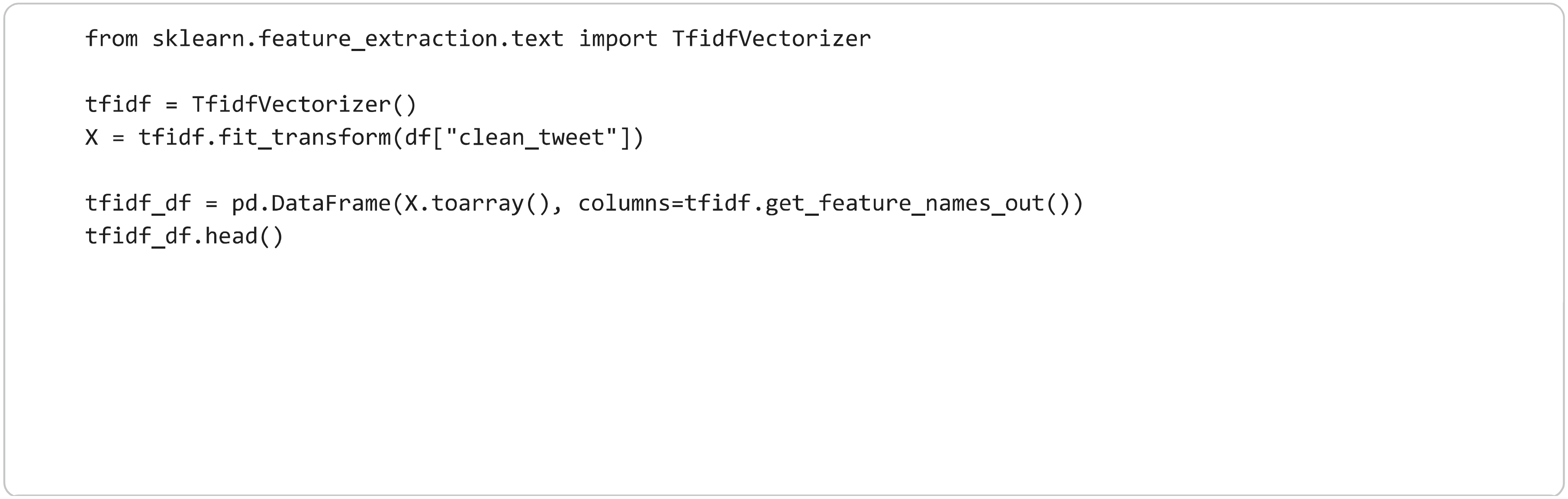


```
import re
import nltk
import spacy
from nltk.corpus import stopwords

nltk.download('stopwords')
stop_words = set(stopwords.words('english'))

nlp = spacy.load("en_core_web_sm")

def clean_text(text):
    text = re.sub(r"http\S+", "", text)     # remove URLs
    text = re.sub(r"@\w+", "", text)        # remove mentions
    text = re.sub(r"#\w+", "", text)        # remove hashtags
    doc = nlp(text.lower())
    tokens = [token.text for token in doc
              if token.text.isalpha() and token.text not in stop_words]
    return " ".join(tokens)

df["clean_tweet"] = df["tweet"].apply(clean_text)
df
```

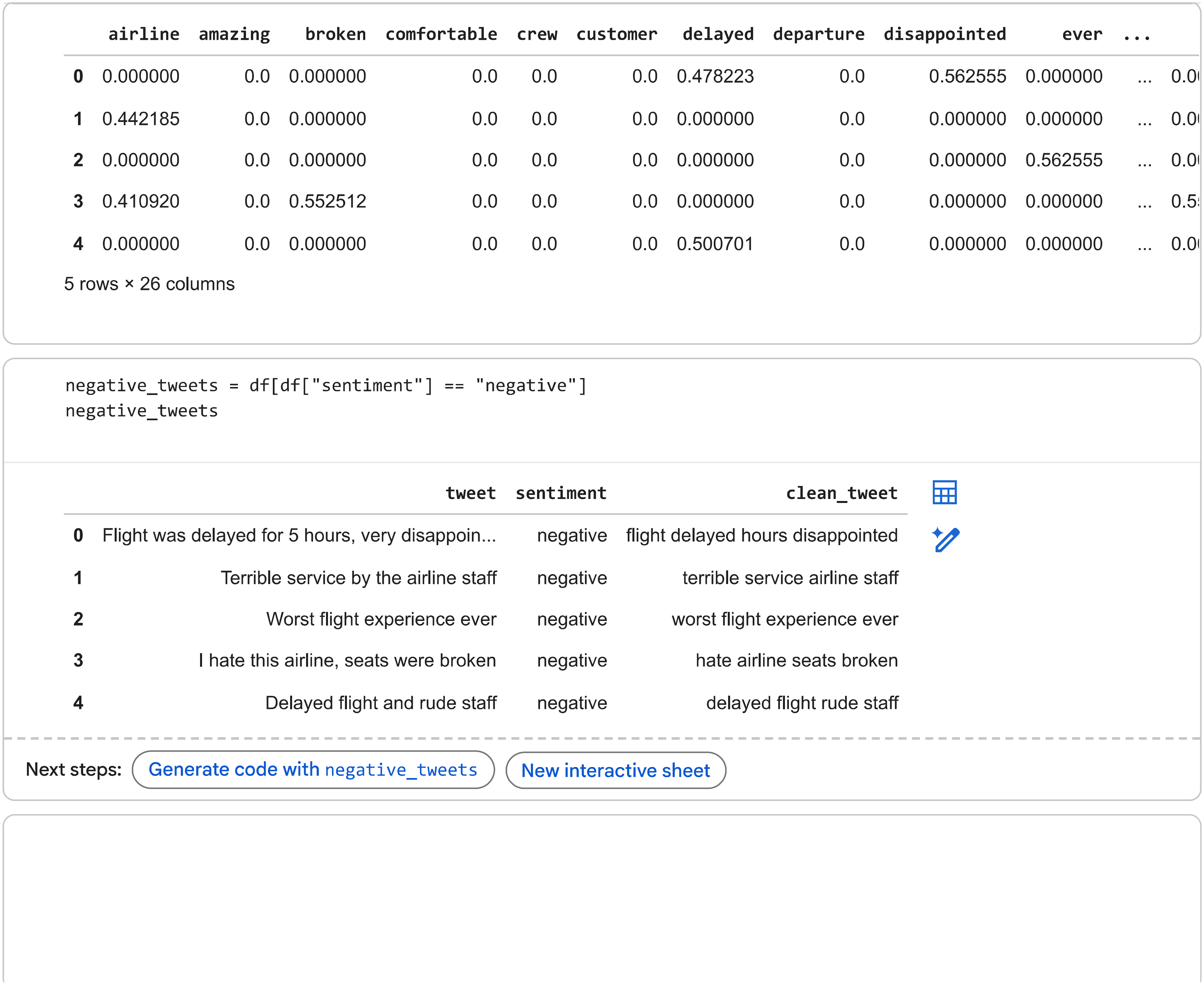


```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

	tweet	sentiment	clean_tweet
0	Flight was delayed for 5 hours, very disappoin...	negative	flight delayed hours disappointed
1	Terrible service by the airline staff	negative	terrible service airline staff
2	Worst flight experience ever	negative	worst flight experience ever
3	I hate this airline, seats were broken	negative	hate airline seats broken
4	Delayed flight and rude staff	negative	delayed flight rude staff
5	Amazing service and friendly crew	positive	amazing service friendly crew
6	Loved the flight experience	positive	loved flight experience
7	Great airline with comfortable seats	positive	great airline comfortable seats
8	Happy with the on-time departure	positive	happy time departure
9	Excellent customer service	positive	excellent customer service



```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf = TfidfVectorizer()
X = tfidf.fit_transform(df["clean_tweet"])

tfidf_df = pd.DataFrame(X.toarray(), columns=tfidf.get_feature_names_out())
tfidf_df.head()
```



```
   airline   amazing  broken    comfortable  crew  customer  delayed   departure  disappointed  ever      ...
0  0.000000  0.0      0.000000  0.0          0.0   0.0       0.478223  0.0        0.562555      0.000000  ...  0.0
1  0.442185  0.0      0.000000  0.0          0.0   0.0       0.000000  0.0        0.000000      0.000000  ...  0.0
2  0.000000  0.0      0.000000  0.0          0.0   0.0       0.000000  0.0        0.000000      0.562555  ...  0.0
3  0.410920  0.0      0.552512  0.0          0.0   0.0       0.000000  0.0        0.000000      0.000000  ...  0.5
4  0.000000  0.0      0.000000  0.0          0.0   0.0       0.500701  0.0        0.000000      0.000000  ...  0.0

5 rows × 26 columns
```

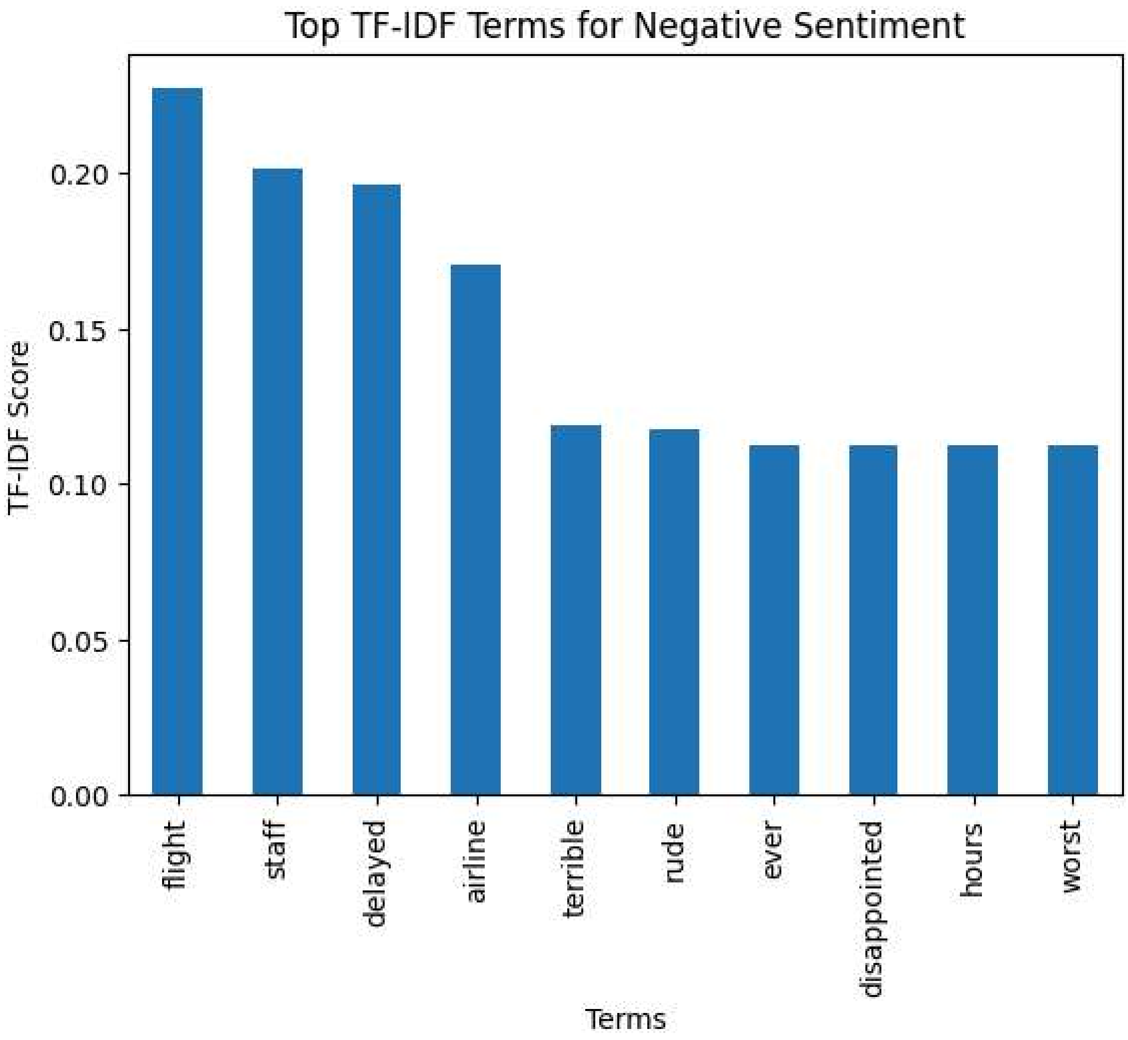
```python
negative_tweets = df[df["sentiment"] == "negative"]
negative_tweets
```

```
                                               tweet  sentiment                      clean_tweet
0  Flight was delayed for 5 hours, very disappoin...  negative   flight delayed hours disappointed
1                 Terrible service by the airline staff  negative   terrible service airline staff
2                        Worst flight experience ever  negative   worst flight experience ever
3                  I hate this airline, seats were broken  negative   hate airline seats broken
4                        Delayed flight and rude staff  negative   delayed flight rude staff
```

```
neg_indices = negative_tweets.index
neg_tfidf = tfidf_df.iloc[neg_indices]

top_terms = neg_tfidf.mean().sort_values(ascending=False).head(10)
top_terms
```

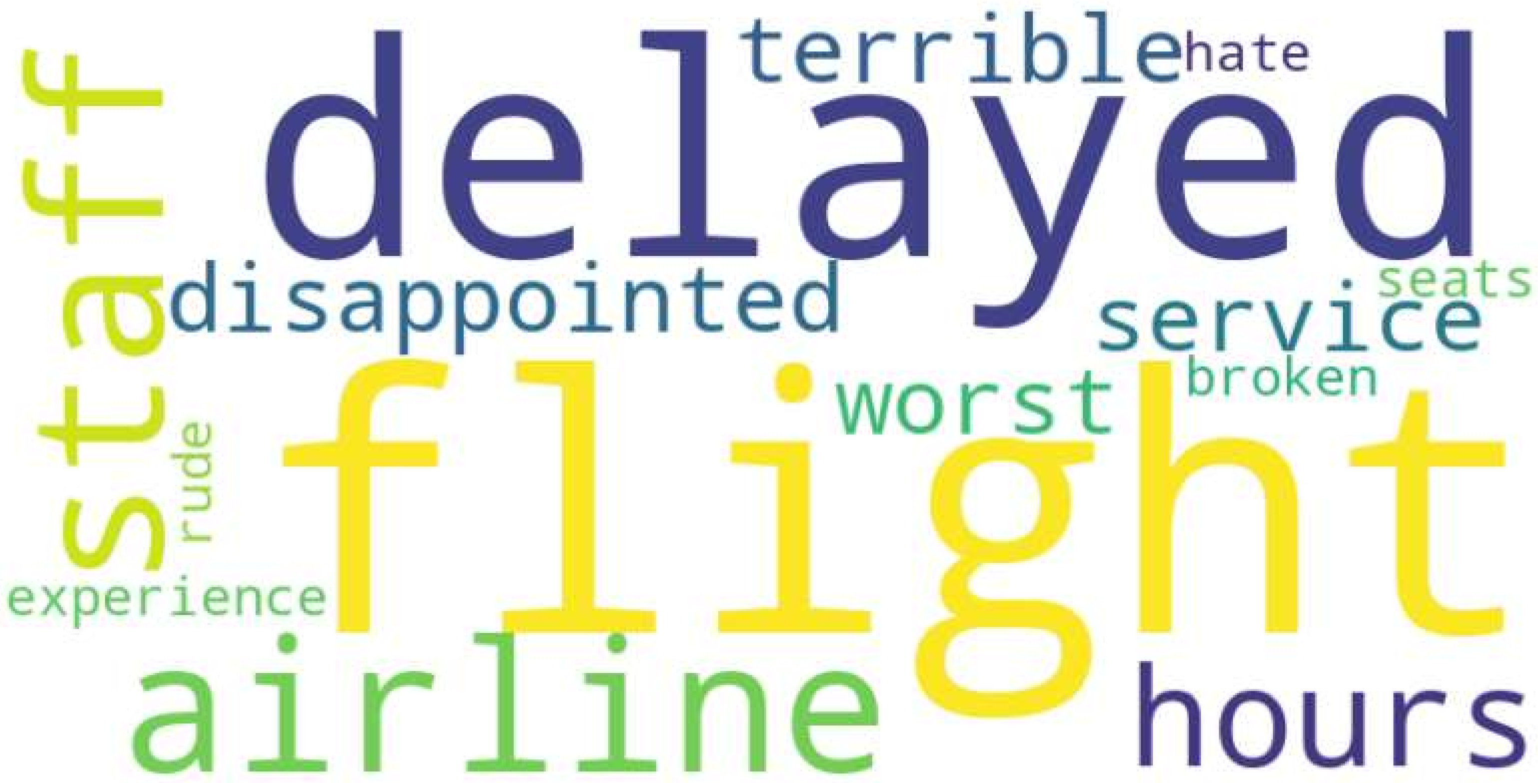
```
                     0
flight        0.226683
staff         0.201225
delayed       0.195785
airline       0.170621
terrible      0.118910
rude          0.117799
ever          0.112511
disappointed  0.112511
hours         0.112511
worst         0.112511

dtype: float64
```

```
import matplotlib.pyplot as plt

top_terms.plot(kind='bar')
plt.title("Top TF-IDF Terms for Negative Sentiment")
plt.xlabel("Terms")
plt.ylabel("TF-IDF Score")
plt.show()
```



```
from wordcloud import WordCloud

wordcloud = WordCloud(
    background_color='white',
    width=800,
    height=400
).generate(" ".join(negative_tweets["clean_tweet"]))

plt.figure(figsize=(10,5))
```

```
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis("off")
    plt.show()
```


delayed
terrible
hate
staff
disappointed
service
seats
broken
worst
flight
rude
experience
airline
hours