```
import pandas as pd
data={'text':
 ["i love this product!",
  "this is the worst",
  "its was ok not great",
  "absoulutly amazing",
   "i didnt like",
    "just average"],
  'sentiment':
       ["positive",
        "negative",
        "neutral",
        "positive",
        "negative",
        "neutral"
                  ]}
df=pd.DataFrame(data)
df.head()
```

	text	sentiment
0	i love this product!	positive
1	this is the worst	negative
2	its was ok not great	neutral
3	absoulutly amazing	positive
4	i didnt like	negative

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
import nltk
import re
nltk.download('stopwords')
from nltk.corpus import stopwords

def clean_text(text):
    text = text.lower()  # lowercase
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    text = text.strip()
    text = ' '.join([word for word in text.split() if word not in stopwords.words('english')])
    return text


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

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

	text	clean_text
0	i love this product!	love product
1	this is the worst	worst
2	its was ok not great	ok great
3	absoulutly amazing	absoulutly amazing
4	i didnt like	didnt like
5	just average	average

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
df['label'] = le.fit_transform(df['sentiment'])
df[['sentiment', 'label']]
```

	sentiment	label
0	positive	2
1	negative	0
2	neutral	1
3	positive	2
4	negative	0
5	neutral	1