```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')


from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import GaussianNB
from sklearn.naive_bayes import BernoulliNB
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import precision_score


dataset=pd.read_csv('/content/emails_2.csv')
dataset.head()
```

	text	spam
0	Subject: naturally irresistible your corporate...	1
1	Subject: the stock trading gunslinger fanny i...	1
2	Subject: unbelievable new homes made easy im ...	1
3	Subject: 4 color printing special request add...	1

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

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5728 entries, 0 to 5727
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   text    5728 non-null   object
 1   spam    5728 non-null   int64 
dtypes: int64(1), object(1)
memory usage: 89.6+ KB
```

```
dataset['spam'].value_counts()
```

spam	count
0	4360
1	1368

```
dataset[dataset['spam']==0]
```

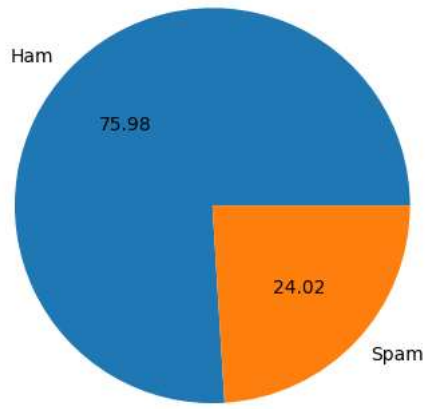
	text	spam
1368	Subject: hello guys , i ' m " bugging you " f...	0
1369	Subject: sacramento weather station fyi - - ...	0
1370	Subject: from the enron india newsdesk - jan 1...	0
1371	Subject: re : powerisk 2001 - your invitation ...	0
1372	Subject: re : resco database and customer capt...	0
...	...	...
5723	Subject: re : research and development charges...	0
5724	Subject: re : receipts from visit jim , than...	0
5725	Subject: re : enron case study update wow ! a...	0
5726	Subject: re : interest david , please , call...	0
5727	Subject: news : aurora 5 . 2 update aurora ve...	0

```
dataset[dataset['spam']==1]
```

	text	spam
0	Subject: naturally irresistible your corporate...	1
1	Subject: the stock trading gunslinger fanny i...	1
2	Subject: unbelievable new homes made easy im ...	1
3	Subject: 4 color printing special request add...	1
4	Subject: do not have money , get software cds ...	1
...	...	...
1363	Subject: are you ready to get it ? hello ! v...	1
1364	Subject: would you like a $ 250 gas card ? do...	1
1365	Subject: immediate reply needed dear sir , i...	1
1366	Subject: wanna see me get fisted ? fist bang...	1
1367	Subject: hot stock info : drgv announces anoth...	1

```
dataset.isnull().sum()
dataset=dataset.drop_duplicates()


plt.pie(dataset['spam'].value_counts(),labels=['Ham','Spam'],autopct='%0.2f')
plt.show()
```



```
import nltk
from nltk.tokenize import word_tokenize
nltk.download('punkt')
```

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

```
dataset['char_num']=dataset['text'].apply(len)
dataset.head(10)
```

	text	spam	char_num
0	Subject: naturally irresistible your corporate...	1	1484
1	Subject: the stock trading gunslinger fanny i...	1	598
2	Subject: unbelievable new homes made easy im ...	1	448
3	Subject: 4 color printing special request add...	1	500
4	Subject: do not have money , get software cds ...	1	235
5	Subject: great nnews hello , welcome to medzo...	1	478
6	Subject: here ' s a hot play in motion homela...	1	9340
7	Subject: save your money buy getting this thin...	1	446
8	Subject: undeliverable : home based business f...	1	507

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

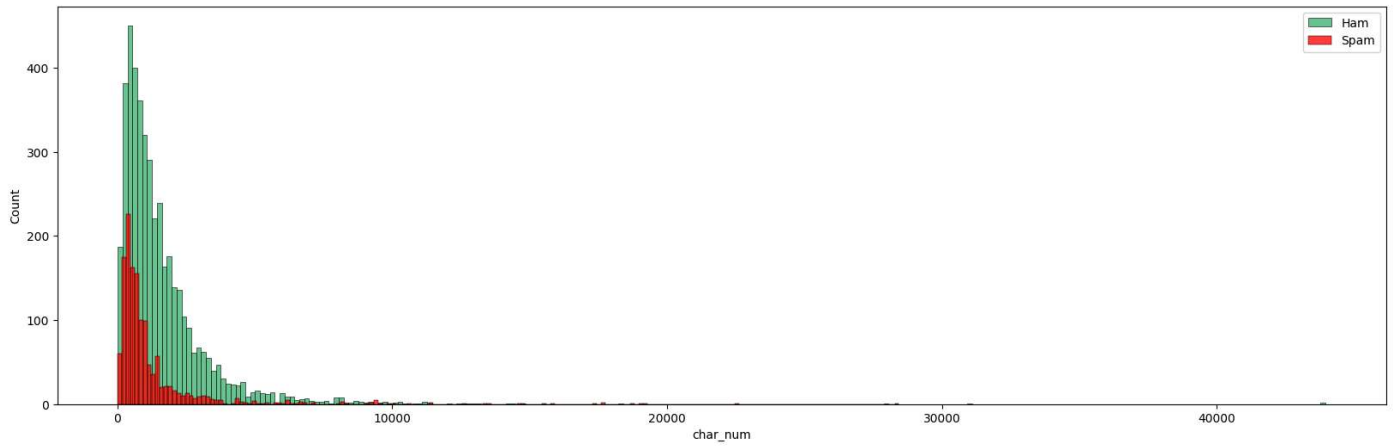
```
dataset['words_num']=dataset['text'].apply(lambda x:len(nltk.word_tokenize(x)))
dataset.head(10)
```

	text	spam	char_num	words_num
0	Subject: naturally irresistible your corporate...	1	1484	325
1	Subject: the stock trading gunslinger fanny i...	1	598	90
2	Subject: unbelievable new homes made easy im ...	1	448	88
3	Subject: 4 color printing special request add...	1	500	99
4	Subject: do not have money , get software cds ...	1	235	53
5	Subject: great nnews hello , welcome to medzo...	1	478	85
6	Subject: here ' s a hot play in motion homela...	1	9340	1704
7	Subject: save your money buy getting this thin...	1	446	97
8	Subject: undeliverable : home based business f...	1	507	122

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

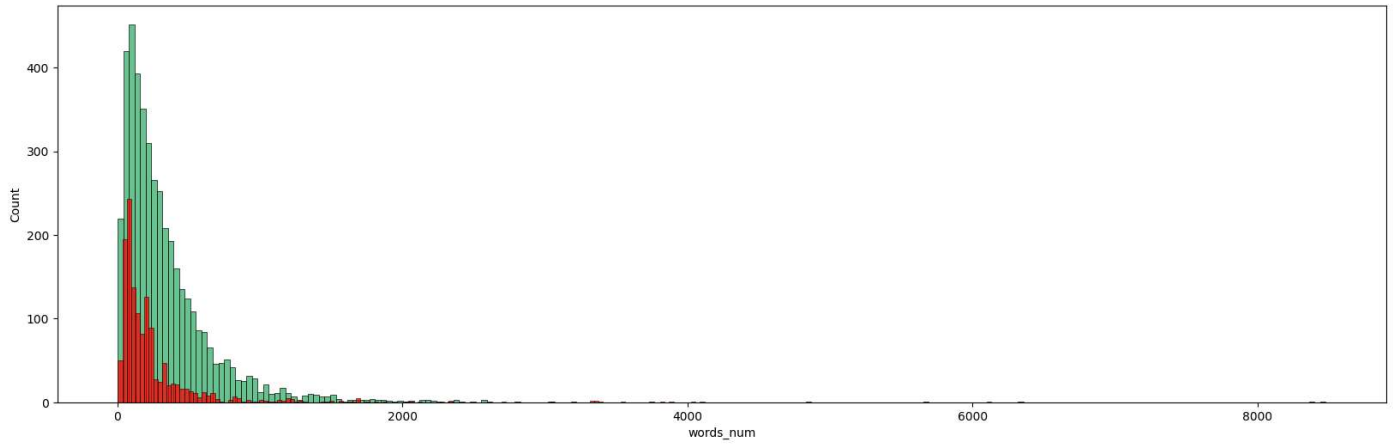
```
plt.figure(figsize=(20,6))
sns.histplot(dataset[dataset['spam']==0]['char_num'],color='mediumseagreen')
sns.histplot(dataset[dataset['spam']==1]['char_num'],color='red')
#plt.legend(['Ham','Spam'])
```

```
<matplotlib.legend.Legend at 0x7cd54e3745b0>
```




```
plt.figure(figsize=(20,6))
sns.histplot(dataset[dataset['spam']==0]['words_num'],color='mediumseagreen')
sns.histplot(dataset[dataset['spam']==1]['words_num'],color='red')
```

```
<Axes: xlabel='words_num', ylabel='Count'>
```




```
vectorizer=CountVectorizer()
X=vectorizer.fit_transform(dataset['text'])
```

```
X_trian,X_test,y_train,y_test=train_test_split(X,dataset['spam'],test_size=0.2)
```

```
model=MultinomialNB()
model.fit(X_trian,y_train)
```

```
MultinomialNB
MultinomialNB()
```

```
y_pred=model.predict(X_test)
acc=accuracy_score(y_test,y_pred)
print(acc)
```

```
0.9920983318700615
```

```
model_2=BernoulliNB()
model_2.fit(X_trian,y_train)
```

```
BernoulliNB
BernoulliNB()
```

```
y_pred1=model_2.predict(X_test)
acc1=accuracy_score(y_test,y_pred1)
print(acc1)
```

```
0.9868305531167691
```

```
model_1=GaussianNB()
model_1.fit(X_trian.toarray(),y_train)
```

```
GaussianNB()
```

```
y_pred2=model_1.predict(X_test.toarray())
acc2=accuracy_score(y_test,y_pred2)
print(acc2)
```

```
0.9604916593503073
```

```
#gaussianNB()
def predictMessage(message):
  messageVector=vectorizer.transform([message])
  prediction=model_1.predict(messageVector.toarray())
  return "Spam" if prediction[0]==1 else "Ham"

usermsg=input("Enter your Message:")
predict=predictMessage(usermsg)
print(f'The message is :{predict}')
```

```
Enter your Message:Well keep in mind I've only got enough gas for one more round trip barring a sudden influx of cash
The message is :Ham
```

```
#MultinomialNb() >>>>>>>>>>>>> combine of gaussianNb() And BernoulliNB()
def predictMessage(message):
  messageVector=vectorizer.transform([message])
  prediction=model.predict(messageVector)
  return "Spam" if prediction[0]==1 else "Ham"

usermsg=input("Enter your Message:")
predict=predictMessage(usermsg)
print(f'The message is :{predict}')
```

```
Enter your Message:Well keep in mind I've only got enough gas for one more round trip barring a sudden influx of cash
The message is :Ham
```

```
#BernoulliNb()
def predictMessage(message):
  messageVector=vectorizer.transform([message])
  prediction=model_2.predict(messageVector.toarray())
  return "Spam" if prediction[0]==1 else "Ham"
```