

```
import numpy as np
import pandas as pd
import seaborn as sns
```

```
df = pd.read_csv('/content/spam (1).csv',encoding='ISO-8859-1')
```

```
df.head(20)
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN
5	spam	FreeMsg Hey there darling it's been 3 week's n...	NaN	NaN	NaN
6	ham	Even my brother is not like to speak with me. ...	NaN	NaN	NaN
7	ham	As per your request 'Melle Melle (Oru Minnamin...	NaN	NaN	NaN
8	spam	WINNER!! As a valued network customer you have...	NaN	NaN	NaN
9	spam	Had your mobile 11 months or more? U R entitle...	NaN	NaN	NaN
10	ham	I'm gonna be home soon and i don't want to tal...	NaN	NaN	NaN
11	spam	SIX chances to win CASH! From 100 to 20,000 po...	NaN	NaN	NaN
12	spam	URGENT! You have won a 1 week FREE membership ...	NaN	NaN	NaN
13	ham	I've been searching for the right words to tha...	NaN	NaN	NaN
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	NaN	NaN	NaN
15	spam	XXXMobileMovieClub: To use your credit, click ...	NaN	NaN	NaN
16	ham	Oh k...i'm watching here:)	NaN	NaN	NaN
17	ham	Eh u remember how 2 spell his name... Yes i di...	NaN	NaN	NaN
18	ham	Fine if thatâs the way u feel. Thatâs the wa...	NaN	NaN	NaN
19	spam	England v Macedonia - dont miss the goals/team...	NaN	NaN	NaN

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0    v1          5572 non-null   object
1    v2          5572 non-null   object
2    Unnamed: 2  50 non-null     object
3    Unnamed: 3  12 non-null     object
4    Unnamed: 4   6 non-null     object
dtypes: object(5)
memory usage: 217.8+ KB
```

```
df.count()
```

```
v1          5572
v2          5572
Unnamed: 2    50
Unnamed: 3    12
Unnamed: 4     6
dtype: int64
```

```
df.shape
```

```
(5572, 5)
```

```
df.dtypes
```

```
v1          object
v2          object
Unnamed: 2   object
```

```
Unnamed: 3    object
Unnamed: 4    object
dtype: object
```

```
df.groupby('v1').describe()
df['spam'] = df['v1'].apply(lambda x:1 if x=='spam' else 0)
df.head()
```

	v1		v2	Unnamed: 2	Unnamed: 3	Unnamed: 4	spam
0	ham	Go until jurong point, crazy.. Available only ...		NaN	NaN	NaN	0
1	ham	Ok lar... Joking wif u oni...		NaN	NaN	NaN	0
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...		NaN	NaN	NaN	1
3	ham	U dun say so early hor... U c already then say...		NaN	NaN	NaN	0
4	ham	Nah I don't think he goes to usf, he lives aro...		NaN	NaN	NaN	0

```
new_df = df[['v1', 'v2', 'spam']]
new_df.head()
```

	v1		v2	spam
0	ham	Go until jurong point, crazy.. Available only ...		0
1	ham	Ok lar... Joking wif u oni...		0
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...		1
3	ham	U dun say so early hor... U c already then say...		0
4	ham	Nah I don't think he goes to usf, he lives aro...		0

```
from sklearn.model_selection import train_test_split as tts
x_train,x_test,y_train,y_test=tts(df.v2,df.spam)
from sklearn.feature_extraction.text import CountVectorizer
v=CountVectorizer()
x_train_count=v.fit_transform(x_train.values)
x_train_count.toarray()[ :2]

array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]])
```

```
from sklearn.naive_bayes import MultinomialNB
model=MultinomialNB()
model.fit(x_train_count,y_train)
```

▼ MultinomialNB

MultinomialNB()

```
emails=["I HAVE A DATE ON SUNDAY WITH WILL!!", "Joking"]
email_count=v.transform(emails)
model.predict(email_count)

array([0, 0])
```

```
x_test_count=v.transform(x_test)
model.score(x_test_count,y_test)

0.9842067480258435
```

