

```
import numpy as np
import pandas as pd

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

df.head()
```

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

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0    Category    5572 non-null   object
1    Message     5572 non-null   object
dtypes: object(2)
memory usage: 87.2+ KB
```

```
df.groupby('Category').describe()
```

Message				
			count	unique
			top	freq
Category				
ham	4825	4516	Sorry, I'll call later	30
spam	747	641	Please call our customer service representativ...	4

```
df['spam'] = df['Category'].apply(lambda x:1 if x=='spam' else 0)

df.head()
```

Category			Message	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

```
new_df = df[['Category', 'Message', 'spam']]

new_df.head()
```

Category			Message	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.Message,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=["Hello sir,do you want a Cricket Match pass?", "Free entry"]
email_count=v.transform(emails)
model.predict(email_count)

array([0, 1])

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

0.9806173725771715
```

