```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB

spam_df = pd.read_csv("spam.csv")

spam_df
```

```
     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...
...       ...                                                ...
5567     spam  This is the 2nd time we have tried 2 contact u...
5568      ham              Will ü b going to esplanade fr home?
5569      ham  Pity, * was in mood for that. So...any other s...
5570      ham  The guy did some bitching but I acted like i'd...
5571      ham                         Rofl. Its true to its name

[5572 rows x 2 columns]
```

```python
spam_df.groupby('Category').describe()
```

```
         Message

           count unique
top
Category

ham         4825   4516                               Sorry, I'll call
later  \
spam         747    641  Please call our customer service
representativ...


          freq
Category
ham         30
spam         4
```

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

spam_df
```

```
     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
...       ...                                                    ...     ...
5567    spam  This is the 2nd time we have tried 2 contact u...       1
5568     ham              Will ü b going to esplanade fr home?       0
5569     ham  Pity, * was in mood for that. So...any other s...       0
5570     ham  The guy did some bitching but I acted like i'd...       0
5571     ham                         Rofl. Its true to its name       0

[5572 rows x 3 columns]

x_train, x_test, y_train, y_test = train_test_split(spam_df.Message,
spam_df.spam, test_size = 0.25)

x_train

2939    Hey i've booked the pilates and yoga lesson al...
2529                       jay says he'll put in  &lt;#&gt;
4969    Future is not what we planned for tomorrow.......
3840    Erm … ill pick you up at about 6.45pm. That'll...
718     Book which lesson? then you msg me... I will c...
                              ...
5546             Ic. There are a lotta childporn cars then.
5437                          Am slow in using biola's fne
1971                                       Nvm it's ok...
4307    Awww dat is sweet! We can think of something t...
487     FREE MESSAGE Activate your 500 FREE Text Messa...
Name: Message, Length: 4179, dtype: object

cv = CountVectorizer()
x_train_count = cv.fit_transform(x_train.values)

x_train_count

<4179x7553 sparse matrix of type '<class 'numpy.int64'>'
    with 55611 stored elements in Compressed Sparse Row format>

model = MultinomialNB()
model.fit(x_train_count, y_train)

MultinomialNB()

email_ham = ["Hey how are you?"]
email_ham_count = cv.transform(email_ham)
model.predict(email_ham_count)

array([0], dtype=int64)
```

```
email_spam = ["Free Gift"]
email_spam_count = cv.transform(email_spam)
model.predict(email_spam_count)
```

```
array([1], dtype=int64)
```

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

```
0.9849246231155779
```