```
# importing the [pandas] for dataset operations
import pandas as pd
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
# The file is zipfile
from zipfile import ZipFile
```

```
with ZipFile('drive/MyDrive/temp.zip','r') as zipObj:
  zipObj.extractall()
```

```
# Loading the dataset and named it as df=DataFrame
df = pd.read_csv('/content/IMDB Dataset.csv.zip')
```

Double-click (or enter) to edit

```
# printing the first 5 rows
df.head()
```

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive

```
# printing the last 5 rows
df.tail()
```

	review	sentiment
49995	i thought this movie did a down right good job...	1
49996	bad plot bad dialogue bad acting idiotic direc...	0
49997	i am a catholic taught in parochial elementary...	0
49998	im going to have to disagree with the previous...	0

```
# printing the columns and it's names
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   review     50000 non-null  object
 1   sentiment  50000 non-null  int64
dtypes: int64(1), object(1)
memory usage: 781.4+ KB
```

```
# checking for the null values
df.isna().sum()
```

	0
review	0
sentiment	0

```
# To determine the No.of rows and columns
df.shape
```

```
(50000, 2)
```

```
# how many are the types of (0 -> review and 1 -> rewiew)
df.value_counts()
```

review	sentiment	count
loved todays show it was a variety and not solely cooking which would have been great too very stimulating and captivating always keeping the viewer peeking around the corner to see what was coming up next she is as down to earth and as personable as you get like one of us which made the show all the more enjoyable special guests who are friends as well made for a nice surprise too loved the first theme and that the audience was invited to play along too i must admit i was shocked to see her come in under her time limits on a few things but she did it and by golly ill be writing those recipes down saving time in the kitchen means more time with family those who havent tuned in yet find out what channel and the time i assure you that you wont be disappointed	1	5
hilarious clean lighthearted and quoteworthy what else can you ask for in a film this is my alltime number one favorite movie ever since i was a little girl ive dreamed of owning a blue van with flames and an observation bubble the cliché characters in ridiculous situations are what make this film such great fun the wonderful comedic chemistry between stephen furst harold and andy tennant melio make up most of my favorite parts of the movie and who didnt love the hopeless awkwardness of flynch dont forget the airport antics of leons cronies dressed up as hari krishnas dancing chanting and playing the tambourineunbeatable the clues are genius the locations are classic and the plot is timeless a word to the wise if you didnt watch this film when you were little it probably wont win a place in your heart today but nevertheless give it a chance you may find that it doesnt matter what you say it doesnt matter what you do youve gotta play	1	4
nickelodeon has gone down the toilet they have kids saying things like oh my god and were screwed this show promotes hate for people who arent good looking or arent in the in crowd it say that sexual promiscuity is alright by having girls slobbering over shirtless boys not to mention the overweight boy who takes off his shirt the main characters basically shun anyone out of the ordinary carlys friend sam who may be a lesbian beats the snot out of anybody that crosses her path which says its alright to be a bch this show has so much negativity in it that nobody should watch it i give it a 0 out of 10	0	3
robert jordan is a television star robert jordan likes things orderly on time and properly executed in his world children are to be seen not heard so why would mr jordan want to become the master of a rambunctious band of boy scouts ratings his staff figures that if learns how to interact with the youth they will be more inclined to watch his show of course watching jordan cope comprises most of the fun like mr belvedere and mr belvedere goes to college this one is sure to please anyone interested in obtaining a copy of this film please write to me at iamaseal2yahoocom	1	3
when i got this movie free from my job along with three other similar movies i watched then with very low expectations now this movie isnt bad per se you get what you pay for it is a tale of love betrayal lies sex scandal everything you want in a movie definitely not a hollywood blockbuster but for cheap thrills it is not that bad i would probably never watch this movie again in a nutshell this is the kind of movie that you would see either very late at night on a local television station that is just wanting to take up some time or you would see it on a sunday afternoon on a local television station that is trying to take up some time despite the bad acting cliché lines and sub par camera work i didnt have the desire to turn off the movie and pretend like it never popped into my dvd player the story has been done many times in many movies this one is no different no better no worse just your average movie	0	3
...	...	...
i picked this dvd up at the dollar store the dvd was on the 2 for 1 rack but since it had michael madsen in it i thought that since i had never seen the movie i bought it anyway i must say that i didnt like the movie the movie played more like a documentary or an advertisement for religion than anything else i found that the directors use of flashbacks did not add to the story line for me i would have preferred to view the story line in chronological order i wont throw it away like one of the other commentators but it may be quite awhile before i would consider watching this movie again who knows since it was michael madsens film debut maybe it might have some archival value at some future date	0	1
i picked this dvd up for 399 at rogers video in order to get enough points to get a better movie for free i never actually was planning on watching this but it started poking at my curiosity and i finally decided to pop in it the dvd player the effects in this movie are horrible and cheap some of the dialog in this movie sounds like it was written by a swear happy 12 year old boy the acting is really cheesy in some parts and the action scenes are completely laughable youll burst out laughing at some parts which was a positive for me because it kept me mildly entertained the plot is some girl has a curse on her which causes her to vomit snakes so some shaman has to get her to los angeles there are also two girls trying to smuggle drugs there and a few other people that are unimportant to the plot not that there really is a plot at alldont expect anything from this movie and dont listen to the cover there are not 100 passengers and 3000 vipers there are 10 passengers and 20 random snakes as for the dvd there is a trailer which is almost as laughable as the film a blooper reel which is just one shot over and over of one actor trying to say train and the deleted scenes are really pointless if they werent good enough to stay in this movie they must be pretty bad	0	1

```
# importing the required libraries
import re
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder

# Text preprocessing function
def preprocess_text(text):
    text = text.lower()
    text = re.sub(r'<br />', ' ', text)  # Remove HTML tags
    text = re.sub(r'[^\w\s]', '', text)  # Remove punctuation
    return text

df['review'] = df['review'].apply(preprocess_text)

# Encode labels
le = LabelEncoder()
df['sentiment'] = le.fit_transform(df['sentiment'])

# Split the dataset into training dataset and testing dataset.
X_train, X_test, y_train, y_test = train_test_split(df['review'], df['sentiment'], test_size=0.2, random_state=42)

```

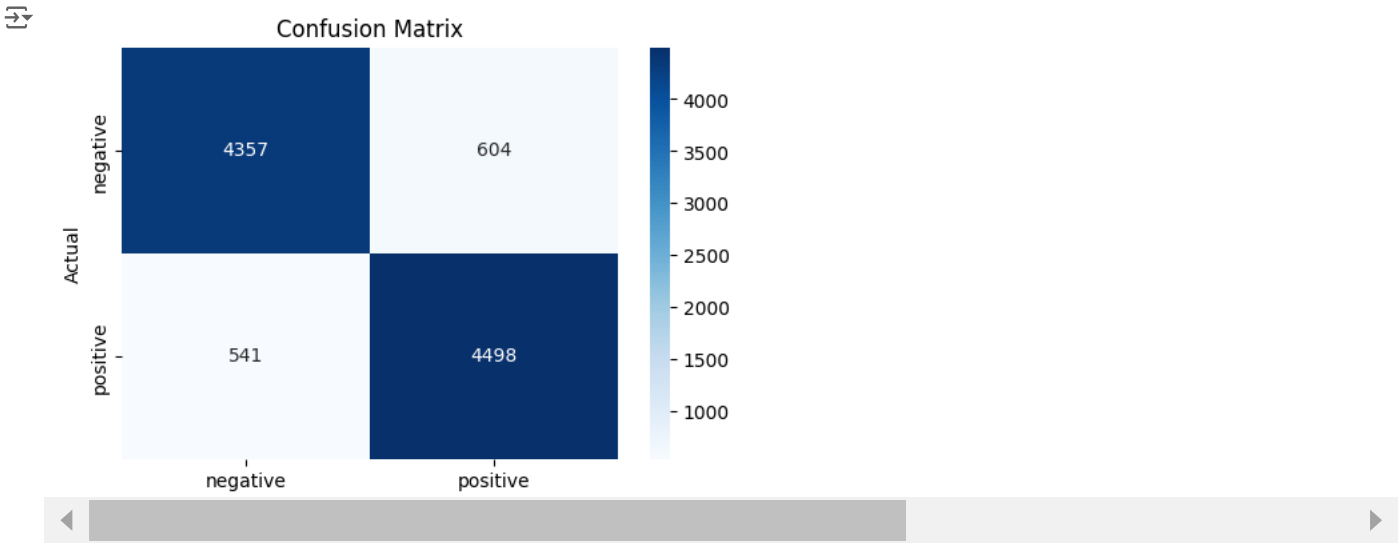
```
# Vectorizing the  text data
vectorizer = CountVectorizer(stop_words='english')
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

```

```
# importing the model and metrics calculation
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

# Train the model
model = LogisticRegression()
model.fit(X_train_vec, y_train)

# Make predictions
y_pred = model.predict(X_test_vec)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy}')
print(classification_report(y_test, y_pred))

```

```
Accuracy: 0.8855
              precision    recall  f1-score   support

           0       0.89      0.88      0.88      4961
           1       0.88      0.89      0.89      5039

    accuracy                           0.89     10000
   macro avg       0.89      0.89      0.89     10000
weighted avg       0.89      0.89      0.89     10000

/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:469: ConvergenceWarning: lbfgs failed to converge (status=
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Default title text

```
# @title Default title text
# plotting the confusion matrix graph by importing the libraries
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix

# Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6,4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=le.classes_, yticklabels=le.classes_)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()

```


Confusion Matrix
Actual
negative
positive
4357
604
541
4498
negative
positive
4000
3500
3000
2500
2000
1500
1000