In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

In [3]:
```
movies.head(1)
```

Out[3]:

	budget	genres	homepage	id	keywords	original_language	original_t
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Av

In [4]:
```
# credits.head(1)['crew'].values #same for the cast a big data set will appear
```

In [5]:
```
credits.head(1)
```

Out[5]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...

In [6]:
```
#merging the 2 datat sets on the title base
movies.merge(credits,on='title').shape #check movis.shape and creditis.shape g
```

Out[6]:
```
(4809, 23)
```

In [7]:
```
movies=movies.merge(credits,on='title')
```

In [8]:
```
movies.head(1)#check the two data sets are merge in last side we have 2 more c
```

Out[8]:

	budget	genres	homepage	id	keywords	original_language	original_t
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Av

1 rows × 23 columns

In [9]:
```
#now let us check it the two data sets are merged or not
#lets check which colums are required for our system we can take out of these
#following colums will take
#genres
#id
#keywords
#title
#overview
#cast
#crew
movies = movies[['movie_id','title','overview','genres','keywords','cast','cre
```

In [10]:
```
movies.info()#here are the only colums which we required
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4809 entries, 0 to 4808
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   movie_id  4809 non-null   int64 
 1   title     4809 non-null   object
 2   overview  4806 non-null   object
 3   genres    4809 non-null   object
 4   keywords  4809 non-null   object
 5   cast      4809 non-null   object
 6   crew      4809 non-null   object
dtypes: int64(1), object(6)
memory usage: 300.6+ KB
```

In [11]:
```
movies.head()
```

Out[11]:

	movie_id	title	overview	genres	keywords	cast	c
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_ "52fe48009251416c750aca "c
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_ "52fe4232c3a36847f800b5 "c
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_ "54805967c3a36829b5002c "c
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_ "52fe4781c3a36847f81398 "c
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_ "52fe479ac3a36847f813ea "c

In [12]:
```
movies.isnull().sum()#now lets check missing data in this for data preprocessi
```

Out[12]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [13]:
```
#here we can see overview of 4 movies is missing so let us drop them
movies.dropna(inplace=True)
```

In [14]:
```
#check again is it drop or not
movies.isnull().sum() #Successfully droped the the missing movies
```

Out[14]:
```
movie_id    0
title       0
overview    0
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [15]:
```
#now let us check for the duplicate row is there or npt in this data set
movies.duplicated().sum()#therefore none of the colums is dublicated
```

Out[15]:
```
0
```

In [16]:
```
movies.iloc[0].genres
```

Out[16]:
```
'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14,
"name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

In [17]:
```
#'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "
# as the data is in this form and we dont want this in that form we want in th
#['action','adventure','fantasy','Scifi']
```

In [18]:
```
def convert(obj):
    L=[]
    for i in ast.literal_eval(obj):
        L.append(i['name'])
    return L
```

In [19]:
```
import ast
ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure
```

Out[19]:
```
[{'id': 28, 'name': 'Action'},
 {'id': 12, 'name': 'Adventure'},
 {'id': 14, 'name': 'Fantasy'},
 {'id': 878, 'name': 'Science Fiction'}]
```

In [20]:
```
movies['genres'].apply(convert)#WE GET THE DATA IN THE FORMAT WE WANT
```

Out[20]:
```
0       [Action, Adventure, Fantasy, Science Fiction]
1                      [Adventure, Fantasy, Action]
2                        [Action, Adventure, Crime]
3                    [Action, Crime, Drama, Thriller]
4                 [Action, Adventure, Science Fiction]
                              ...
4804                       [Action, Crime, Thriller]
4805                               [Comedy, Romance]
4806                [Comedy, Drama, Romance, TV Movie]
4807                                               []
4808                                     [Documentary]
Name: genres, Length: 4806, dtype: object
```

In [21]:
```
movies['genres'] = movies['genres'].apply(convert)
```

In [22]:
```
movies.head()
```

Out[22]:

	movie_id	title	overview	genres	keywords	cast	cre
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_ic "52fe48009251416c750aca23 "de
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_ic "52fe4232c3a36847f800b579 "de
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_ic "54805967c3a36829b5002c41 "de
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_ic "52fe4781c3a36847f81398c3 "de
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_ic "52fe479ac3a36847f813eaa3 "de

In [23]:
```
movies['keywords'].apply(convert)
```

Out[23]:
```
0       [culture clash, future, space war, space colon...
1       [ocean, drug abuse, exotic island, east india ...
2       [spy, based on novel, secret agent, sequel, mi...
3       [dc comics, crime fighter, terrorist, secret i...
4       [based on novel, mars, medallion, space travel...
                              ...
4804    [united states-mexico barrier, legs, arms, pap...
4805                                                   []
4806    [date, love at first sight, narration, investi...
4807                                                   []
4808              [obsession, camcorder, crush, dream girl]
Name: keywords, Length: 4806, dtype: object
```

In [24]:
```
movies['keywords'] = movies['keywords'].apply(convert)
```

In [25]:
```
movies.head()
```

Out[25]:

	movie_id	title	overview	genres	keywords	cast	cre
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_ic "52fe48009251416c750aca23 "de
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_ic "52fe4232c3a36847f800b579 "de
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_ic "54805967c3a36829b5002c41 "de
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_ic "52fe4781c3a36847f81398c3 "de
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_ic "52fe479ac3a36847f813eaa3 "de

In [26]:
```
def convert3(obj):
    L=[]
    counter = 0
    for i in ast.literal_eval(obj):
        if counter != 3:
            L.append(i['name'])
            counter +=1
        else:
            break
    return L
```

In [27]:
```
movies['cast'].apply(convert3)#now here we only get 3 actor/actress name  whic
```

Out[27]:
```
0         [Sam Worthington, Zoe Saldana, Sigourney Weaver]
1            [Johnny Depp, Orlando Bloom, Keira Knightley]
2             [Daniel Craig, Christoph Waltz, Léa Seydoux]
3             [Christian Bale, Michael Caine, Gary Oldman]
4           [Taylor Kitsch, Lynn Collins, Samantha Morton]
                               ...
4804    [Carlos Gallardo, Jaime de Hoyos, Peter Marqua...
4805          [Edward Burns, Kerry Bishé, Marsha Dietlein]
4806            [Eric Mabius, Kristin Booth, Crystal Lowe]
4807             [Daniel Henney, Eliza Coupe, Bill Paxton]
4808    [Drew Barrymore, Brian Herzlinger, Corey Feldman]
Name: cast, Length: 4806, dtype: object
```

In [28]:
```
movies['cast'] = movies['cast'].apply(convert3)
```

In [29]:
```
movies.head()
```

Out[29]:

	movie_id	title	overview	genres	keywords	cast	cr
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[{"credit_ "52fe48009251416c750aca2 "d
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[{"credit_ "52fe4232c3a36847f800b57 "d
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[{"credit_ "54805967c3a36829b5002c4 "d
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[{"credit_ "52fe4781c3a36847f81398c "d
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[{"credit_ "52fe479ac3a36847f813eaa "d

In [30]:
```
movies['crew'][0]
```

Out[30]:
```
'[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gend
er": 0, "id": 1721, "job": "Editor", "name": "Stephen E. Rivkin"}, {"credi
t_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id":
496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54
491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "j
ob": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491c
b70e0a267480001bd0", "department": "Sound", "gender": 0, "id": 900, "job":
"Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "5
39c4a4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1
262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "5544ee3b925141
499f0008fc", "department": "Sound", "gender": 2, "id": 1729, "job": "Origi
nal Music Composer", "name": "James Horner"}, {"credit_id": "52fe480092514
16c750ac9c3", "department": "Directing", "gender": 2, "id": 2710, "job":
"Director", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac
9d9", "department": "Writing", "gender": 2, "id": 2710, "job": "Writer",
"name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "depar
tment": "Editing", "gender": 2, "id": 2710, "job": "Editor", "name": "Jame
s Cameron"}, {"credit_id": "52fe48009251416c750aca29", "department": "Prod
uction", "gender": 2, "id": 2710, "job": "Producer", "name": "James Camero
```

In [31]:
```
def fetch_director(obj):
    L = []
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            L.append(i['name'])
            break
    return L
```

In [32]:
```
movies['crew'].apply(fetch_director)#now here we only get the director of the
```

Out[32]:
```
0           [James Cameron]
1          [Gore Verbinski]
2              [Sam Mendes]
3       [Christopher Nolan]
4          [Andrew Stanton]
                 ...
4804     [Robert Rodriguez]
4805         [Edward Burns]
4806          [Scott Smith]
4807          [Daniel Hsia]
4808     [Brian Herzlinger]
Name: crew, Length: 4806, dtype: object
```

In [33]:
```
movies ['crew'] = movies['crew'].apply(fetch_director)
```

In [34]:
```
movies.head()
```

Out[34]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

In [35]:
```
movies['overview'][0]#this is a striing and i will convert it into a list to c
```

Out[35]:
```
'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora o
n a unique mission, but becomes torn between following orders and protecting
an alien civilization.'
```

In [36]:
```
movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

In [37]:
```
movies.head()
```

Out[37]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

In [38]:
```
#we will remove the spaces in between the words 'sam worthington to samwordint
```

In [39]:
```
movies['genres']=movies['genres'].apply(lambda x:[i.replace(" ","") for i in x
movies['keywords']=movies['keywords'].apply(lambda x:[i.replace(" ","") for i
movies['cast']=movies['cast'].apply(lambda x:[i.replace(" ","") for i in x])
movies['crew']=movies['crew'].apply(lambda x:[i.replace(" ","") for i in x])
```

In [40]:
```
movies.head()
```

Out[40]:

	movie_id	title	overview	genres	keywords	cast	
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[culturecclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCam
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbi
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMer
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherN
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewSta

In [41]:
```
movies['tags'] = movies['overview'] + movies['cast'] + movies['crew'] + movies
```

In [42]:
```
movies.head()
```

Out[42]:

	movie_id	title	overview	genres	keywords	cast	
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[culturecclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCam
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbi
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMer
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherN
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewSta

In [43]:
```
new_df = movies[['movie_id','title','tags']]
```

In [44]:
```
new_df
```

Out[44]:

	movie_id	title	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...
2	206647	Spectre	[A, cryptic, message, from, Bond’s, past, send...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...
...	...	...	...
4804	9367	El Mariachi	[El, Mariachi, just, wants, to, play, his, gui...
4805	72766	Newlyweds	[A, newlywed, couple's, honeymoon, is, upended...
4806	231617	Signed, Sealed, Delivered	["Signed,, Sealed,, Delivered", introduces, a,...
4807	126186	Shanghai Calling	[When, ambitious, New, York, attorney, Sam, is...
4808	25975	My Date with Drew	[Ever, since, the, second, grade, when, he, fi...

4806 rows × 3 columns

In [45]:
```
new_df['tags'].apply(lambda x:" ".join(x))#will convert list into strings
```

Out[45]:
```
0       In the 22nd century, a paraplegic Marine is di...
1       Captain Barbossa, long believed to be dead, ha...
2       A cryptic message from Bond’s past sends him o...
3       Following the death of District Attorney Harve...
4       John Carter is a war-weary, former military ca...
                              ...
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: tags, Length: 4806, dtype: object
```

In [46]:
```
new_df['tags'] = new_df['tags'].apply(lambda x:" ".join(x))
```

```
C:\Users\Yash\AppData\Local\Temp\ipykernel_352\3089450492.py:1: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  new_df['tags'] = new_df['tags'].apply(lambda x:" ".join(x))
```

In [47]:
```
new_df.head()
```

Out[47]:

	movie_id	title	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	A cryptic message from Bond’s past sends him o...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...
4	49529	John Carter	John Carter is a war-weary, former military ca...

In [48]:
```
!pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\yash\anaconda3\lib\site-packa
ges (3.8.1)
Requirement already satisfied: click in c:\users\yash\anaconda3\lib\site-pack
ages (from nltk) (8.0.4)
Requirement already satisfied: joblib in c:\users\yash\anaconda3\lib\site-pac
kages (from nltk) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\yash\anaconda3\lib
\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\yash\anaconda3\lib\site-packa
ges (from nltk) (4.65.0)
Requirement already satisfied: colorama in c:\users\yash\anaconda3\lib\site-p
ackages (from click->nltk) (0.4.6)
```

In [49]:
```
import nltk
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
```

In [50]:
```
def stem(text):
    y = []

    for i in text.split():
        y.append(ps.stem(i))

    return " ".join(y)
```

In [51]:
```
new_df['tags'] = new_df['tags'].apply(stem)
```

```
C:\Users\Yash\AppData\Local\Temp\ipykernel_352\3213734980.py:1: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  new_df['tags'] = new_df['tags'].apply(stem)
```

In [52]:
```
new_df['tags'][0]
```

Out[52]:
```
'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a u
niqu mission, but becom torn between follow order and protect an alien civili
zation. samworthington zoesaldana sigourneyweav jamescameron action adventur
fantasi sciencefict cultureclash futur spacewar spacecoloni societi spacetrav
el futurist romanc space alien tribe alienplanet cgi marin soldier battl love
affair antiwar powerrel mindandsoul 3d'
```

In [ ]:
```

```

In [53]:
```
new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

```
C:\Users\Yash\AppData\Local\Temp\ipykernel_352\3214958533.py:1: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

In [54]:
```
new_df.head()
```

Out[54]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond’ past send him on a...
3	49026	The Dark Knight Rises	follow the death of district attorney harvey d...
4	49529	John Carter	john carter is a war-weary, former militari ca...

In [55]:
```
#here vectorization starts
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,stop_words='english')
```

In [56]:
```
vectors = cv.fit_transform(new_df['tags']).toarray().shape #4806 movies done i
```

In [57]:
```
vectors = cv.fit_transform(new_df['tags']).toarray()
```

In [58]:
```
vectors #now each and every movie is in the form of vectors
```

Out[58]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [59]:
```
vectors[0]
```

Out[59]:
```
array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

In [60]:
```
len(cv.get_feature_names_out())#these 5000 word are similar like action and ac
```

Out[60]:
```
5000
```

In [61]:
```
#for applying stemping we will install nltk library
ps.stem('loved','loving')
```

Out[61]:
```
'love'
```

In [62]:
```
stem('In the 22nd century, a paraplegic Marine is dispatched to the moon Pando
```

Out[62]:
```
'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a u
niqu mission, but becom torn between follow order and protect an alien civili
zation. samworthington zoesaldana sigourneyweav jamescameron action adventur
fantasi sciencefict cultureclash futur spacewar spacecoloni societi spacetrav
el futurist romanc space alien tribe alienplanet cgi marin soldier battl love
affair antiwar powerrel mindandsoul 3d'
```

In [63]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

In [64]:
```
similarity = cosine_similarity(vectors)#taking each distance of each movie fro
```

In [65]:
```
sorted(list(enumerate(similarity[0])),reverse = True,key=lambda x:x[1])[1:6]#c
```

Out[65]:
```
[(1216, 0.28676966733820225),
 (2409, 0.26901379342448517),
 (3730, 0.2605130246476754),
 (507, 0.255608593705383),
 (539, 0.2503866978335957)]
```

In [66]:
```
def recommend(movie):
    movie_index =new_df[new_df['title'] == movie].index[0]
    distances = similarity[movie_index]
    movies_list = sorted(list(enumerate(distances)),reverse = True,key=lambda

    for i in movies_list:
        print(new_df.iloc[i[0]].title)
```

In [75]:
```
recommend('Avatar')
```

```
Aliens vs Predator: Requiem
Aliens
Falcon Rising
Independence Day
Titan A.E.
```

In [76]:
```
recommend('Batman Begins')
```

```
The Dark Knight
Batman
Batman
The Dark Knight Rises
10th & Wolf
```

In [69]:
```
#FINAALY MODEL IS BUILT NOW WE WILL CONVERNT IT INTO AN WEBSITE.
```

In [70]:
```
import pickle
```

In [71]:
```
pickle.dump(new_df,open('movies.pkl','wb'))
```

In [72]:
```
new_df['title'].values
```

Out[72]:
```
array(['Avatar', "Pirates of the Caribbean: At World's End", 'Spectre',
       ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
       'My Date with Drew'], dtype=object)
```

In [73]:
```
#as it is giving error so we will pass dicticnory rather then that
pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

In [74]:
```
pickle.dump(similarity,open('similarity.pkl','wb'))
```

In [ ]:
```

```