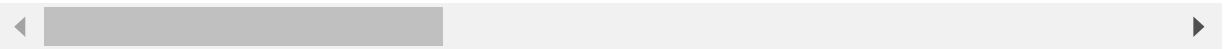


```
import pandas as pd
import numpy as np
```

```
movies=pd.read_csv("C:/Users/zoaah/Downloads/archive/tmdb_5000_movies.csv")
credits=pd.read_csv("C:/Users/zoaah/Downloads/archive/tmdb_5000_credits.csv")
```

```
movies.head(3)
```

	budget	genres	homepage	id	keywords	original_language
0	237000000	{{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	{{"id": 1463, "name": "culture clash"}, {"id": "...	en
1	300000000	{{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	{{"id": 270, "name": "ocean"}, {"id": 726, "na...	en
2	245000000	{{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	{{"id": 470, "name": "spy"}, {"id": 818, "name...	en



```
credits.head(1)
```

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

In [5]:

```
credits.head(1)['cast']
```

Out[5]:

```
0    [{"cast_id": 242, "character": "Jake Sully", "...  
Name: cast, dtype: object
```

In [6]:

```
credits.head(1)['cast'].values
```

Out[6]:

```
array([{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a368  
5532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington", "order": 0},  
{ "cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251416c750ac9cb",  
"gender": 1, "id": 8691, "name": "Zoe Saldana", "order": 1}, {"cast_id": 25, "cha  
racter": "Dr. Grace Augustine", "credit_id": "52fe48009251416c750aca39", "gende  
r": 1, "id": 10205, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "char  
acter": "Col. Quaritch", "credit_id": "52fe48009251416c750ac9cf", "gender": 2, "i  
d": 32747, "name": "Stephen Lang", "order": 3}, {"cast_id": 5, "character": "Trud  
y Chacon", "credit_id": "52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "na  
me": "Michelle Rodriguez", "order": 4}, {"cast_id": 8, "character": "Selfridge",  
"credit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovan  
ni Ribisi", "order": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_i  
d": "52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "name": "Joel David Moo  
re", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_id": "52fe480092514  
16c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder", "order": 7}, {"cas  
t_id": 11, "character": "Eytukan", "credit_id": "52fe48009251416c750ac9ed", "gend  
er": 2, "id": 15853, "name": "Wes Studi", "order": 8}, {"cast_id": 10, "characte  
r": "Tsu'utev". "credit_id": "52fe48009251416c750ac9e9". "gender": 2. "id": 1096
```

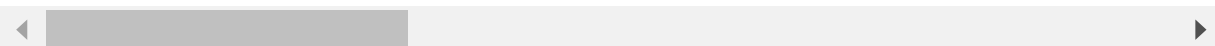
In [7]:

```
#merging of two Dataframes on the basis of title  
movies.merge(credits,on='title')
```

Out[7]:

	budget	genres	homepage	id	keywords
0	237000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}], [{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}], [{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]	http://www.avatarmovie.com/	19995	[[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}], [{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]
1	300000000	[[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}], [{"id": 14, "name": "Action"}], [{"id": 14, "name": "Action"}]	http://disney.go.com/disneypictures/pirates/	285	[[{"id": 270, "name": "ocean"}, {"id": 726, "name": "ocean"}], [{"id": 270, "name": "ocean"}, {"id": 726, "name": "ocean"}]
2	245000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}]	http://www.sonypictures.com/movies/spectre/	206647	[[{"id": 470, "name": "spy"}, {"id": 818, "name": "spy"}], [{"id": 470, "name": "spy"}, {"id": 818, "name": "spy"}]
3	250000000	[[{"id": 28, "name": "Action"}, {"id": 80, "name": "Action"}], [{"id": 80, "name": "Action"}], [{"id": 80, "name": "Action"}]	http://www.thedarkknighttrises.com/	49026	[[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "dc comics"}], [{"id": 849, "name": "dc comics"}, {"id": 853, "name": "dc comics"}]
4	260000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}]	http://movies.disney.com/john-carter	49529	[[{"id": 818, "name": "based on novel"}, {"id": 819, "name": "based on novel"}], [{"id": 818, "name": "based on novel"}, {"id": 819, "name": "based on novel"}]
...
4804	220000	[[{"id": 28, "name": "Action"}, {"id": 80, "name": "Action"}], [{"id": 80, "name": "Action"}], [{"id": 80, "name": "Action"}]	NaN	9367	[[{"id": 5616, "name": "united states\u2013mexico"}, {"id": 5617, "name": "united states\u2013mexico"}], [{"id": 5616, "name": "united states\u2013mexico"}, {"id": 5617, "name": "united states\u2013mexico"}]
4805	9000	[[{"id": 35, "name": "Comedy"}, {"id": 10749, "name": "Comedy"}], [{"id": 10749, "name": "Comedy"}], [{"id": 10749, "name": "Comedy"}]	NaN	72766	[[{"id": 10749, "name": "Comedy"}, {"id": 10750, "name": "Comedy"}], [{"id": 10749, "name": "Comedy"}, {"id": 10750, "name": "Comedy"}]
4806	0	[[{"id": 35, "name": "Comedy"}, {"id": 18, "name": "Comedy"}], [{"id": 18, "name": "Comedy"}], [{"id": 18, "name": "Comedy"}]	http://www.hallmarkchannel.com/signedsealeddelivered/	231617	[[{"id": 248, "name": "date"}, {"id": 699, "name": "date"}], [{"id": 248, "name": "date"}, {"id": 699, "name": "date"}]
4807	0	[[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}], [{"id": 14, "name": "Action"}], [{"id": 14, "name": "Action"}]	http://shanghaicalling.com/	126186	[[{"id": 126186, "name": "shanghai"}, {"id": 126187, "name": "shanghai"}], [{"id": 126186, "name": "shanghai"}, {"id": 126187, "name": "shanghai"}]
4808	0	[[{"id": 99, "name": "Documentary"}], [{"id": 99, "name": "Documentary"}], [{"id": 99, "name": "Documentary"}]	NaN	25975	[[{"id": 1523, "name": "obsession"}, {"id": 224, "name": "obsession"}], [{"id": 1523, "name": "obsession"}, {"id": 224, "name": "obsession"}]

4809 rows × 23 columns



In [8]:

```
movies.merge(credits,on='title') .shape
```

Out[8]:

(4809, 23)

In [9]:

```
movies.shape
```

Out[9]:

(4803, 20)

In [10]:

```
credits.shape
```

Out[10]:

(4803, 4)

In [11]:

```
#new dataframe of movies
movies=movies.merge(credits,on='title')
```

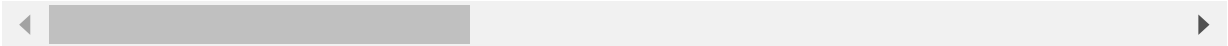
In [12]:

```
movies.head(1)
```

Out[12]:

	budget	genres	homepage	id	keywords	original_language	original_title	over
0	237000000	[[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":	en	Avatar	cer para M...

1 rows × 23 columns



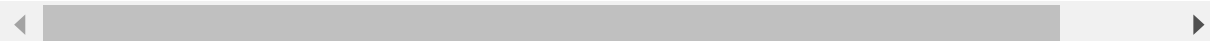
In [13]:

```
#keeping certain columns and removing the columns that are not necessary for our analysis  
#genres  
#id  
#title  
#keywords  
#overview  
#cast  
#crew  
movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]
```

Out[13]:

	movie_id	title	overview	genres	keywords	cast
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", "...
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...
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...
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...
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...
...
4804	9367	El Mariachi	El Mariachi just wants to play his guitar and ...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 5616, "name": "united states\u2013mexi...	[{"cast_id": 1, "character": "El Mariachi", "c...
4805	72766	Newlyweds	A newlywed couple's honeymoon is upended by th...	[{"id": 35, "name": "Comedy"}, {"id": 10749, "...	[]	[{"cast_id": 1, "character": "Buzzy", "credit_...
4806	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...	[{"id": 35, "name": "Comedy"}, {"id": 18, "nam...	[{"id": 248, "name": "date"}, {"id": 699, "nam...	[{"cast_id": 8, "character": "Oliver O\u201919To...
4807	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...	[]	[]	[{"cast_id": 3, "character": "Sam", "credit_id...
4808	25975	My Date with Drew	Ever since the second grade when he first saw ...	[{"id": 99, "name": "Documentary"}]	[{"id": 1523, "name": "obsession"}, {"id": 224...	[{"cast_id": 3, "character": "Herself", "credi...

4809 rows × 7 columns



In [14]:

```
# final new dataframe for the recommender system
movies=movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```

In [15]:

```
movies.head()
```

Out[15]:

	movie_id	title	overview	genres	keywords	cast	crew
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_id": "52fe48009251416c750aca23", "de...
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_id": "52fe4232c3a36847f800b579", "de...
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_id": "54805967c3a36829b5002c41", "de...
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_id": "52fe4781c3a36847f81398c3", "de...
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_id": "52fe479ac3a36847f81398c3", "de...

In [16]:

```
#detecting missing values
movies.isnull().sum()
```

Out[16]:

```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```


In [17]:

```
movies.dropna(inplace=True)
```

In [18]:

```
#removed the movies that had missing values in overview column  
movies.isnull().sum()
```

Out[18]:

```
movie_id    0  
title       0  
overview    0  
genres      0  
keywords    0  
cast        0  
crew        0  
dtype: int64
```

In [19]:

```
#to check for duplicate data  
movies.duplicated().sum()
```

Out[19]:

```
0
```

In [20]:

```
#reformatting of columns  
#here the format is not properly shown  
movies.iloc[0].genres
```

Out[20]:

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

In [21]:

```
#converting string of list into list  
import ast  
ast.literal_eval('{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name":  
'
```

Out[21]:

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

In [22]:

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

In [23]:

```
movies['genres'].apply(convert)
```

Out[23]:

```
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 [24]:

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

In [25]:

```
movies.head()
```

Out[25]:

	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]	{{"id": 1463, "name": "culture clash"}, {"id":...	{{"cast_id": 242, "character": "Jake Sully", "...	{{"credit_id": "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_id": "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_id": "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_id": "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_id": "52fe479ac3a36847f813eaa3", "de...

In [26]:

```
movies['keywords'].apply(convert)
```

Out[26]:

```
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 [27]:

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

In [28]:

```
movies.head()
```

Out[28]:

	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...	{ "cast_id": 242, "character": "Jake Sully", "...	{ "credit_id": "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_id": "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_id": "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_id": "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_id": "52fe479ac3a36847f8139aaa3", "de...

In [29]:

```
#cast column for first movie
movies['cast'][0]
```

Out[29]:

```
{ "cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington", "order": 0}, { "cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana", "order": 1}, { "cast_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52fe48009251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney Weaver", "order": 2}, { "cast_id": 4, "character": "Col. Quaritch", "credit_id": "52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang", "order": 3}, { "cast_id": 5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "name": "Michelle Rodriguez", "order": 4}, { "cast_id": 8, "character": "Selfridge", "credit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovanni Ribisi", "order": 5}, { "cast_id": 7, "character": "Norm Spellman", "credit_id": "52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "name": "Joel David Moore", "order": 6}, { "cast_id": 9, "character": "Moat", "credit_id": "52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder", "order": 7}, { "cast_id": 11, "character": "Eytukan", "credit_id": "52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "name": "Wes Studi", "order": 8}, { "cast_id": 10, "character": "Tsu'tev", "credit_id": "52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Tsu'tev", "order": 9}
```

In [30]:

```
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 [31]:

```
movies['cast'].apply(convert3)
```

Out[31]:

```
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 Mabiús, 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 [32]:

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

In [33]:

```
movies.head()
```

Out[33]:

	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]	[{"credit_id": "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 ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[{"credit_id": "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...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[{"credit_id": "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...	[Christian Bale, Michael Caine, Gary Oldman]	[{"credit_id": "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...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [34]:

```
movies['cast'][0]
```

Out[34]:

```
['Sam Worthington', 'Zoe Saldana', 'Sigourney Weaver']
```

In [35]:

```
#cast column for first movie
movies['crew'][0]
```

Out[35]:

```
{'credit_id': '52fe48009251416c750aca23', 'department': 'Editing', 'gender': 0,
'id': 1721, 'job': 'Editor', 'name': 'Stephen E. Rivkin'}, {'credit_id': '539c47e
cc3a36810e3001f87', 'department': 'Art', 'gender': 2, 'id': 496, 'job': 'Producti
on Design', 'name': 'Rick Carter'}, {'credit_id': '54491c89c3a3680fb4001cf7', 'de
partment': 'Sound', 'gender': 0, 'id': 900, 'job': 'Sound Designer', 'name': 'Chr
istopher Boyes'}, {'credit_id': '54491cb70e0a267480001bd0', 'department': 'Soun
d', 'gender': 0, 'id': 900, 'job': 'Supervising Sound Editor', 'name': 'Christoph
er Boyes'}, {'credit_id': '539c4a4cc3a36810c9002101', 'department': 'Production',
'gender': 1, 'id': 1262, 'job': 'Casting', 'name': 'Mali Finn'}, {'credit_id': '5
544ee3b925141499f0008fc', 'department': 'Sound', 'gender': 2, 'id': 1729, 'job':
'Original Music Composer', 'name': 'James Horner'}, {'credit_id': '52fe4800925141
6c750ac9c3', 'department': 'Directing', 'gender': 2, 'id': 2710, 'job': 'Directo
r', 'name': 'James Cameron'}, {'credit_id': '52fe48009251416c750ac9d9', 'departme
nt': 'Writing', 'gender': 2, 'id': 2710, 'job': 'Writer', 'name': 'James Camero
n'}, {'credit_id': '52fe48009251416c750aca17', 'department': 'Editing', 'gender':
2, 'id': 2710, 'job': 'Editor', 'name': 'James Cameron'}, {'credit_id': '52fe4800
9251416c750aca29', 'department': 'Production', 'gender': 2, 'id': 2710, 'job': 'P
roducer', 'name': 'James Cameron'}. {'credit_id': '52fe48009251416c750aca3f', 'de
```

In [36]:

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

In [37]:

```
movies['crew'].apply(fetch_director)
```

Out[37]:

```
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 [38]:

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

In [39]:

```
movies.head()
```

Out[39]:

	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 [40]:

```
movies['overview'][0] #its a string
```

Out[40]:

```
'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization.'
```

In [41]:

```
#to convert string into list
movies['overview'].apply(lambda x:x.split())
```

Out[41]:

```
0    [In, the, 22nd, century,, a, paraplegic, Marin...
1    [Captain, Barbossa,, long, believed, to, be, d...
2    [A, cryptic, message, from, Bond's, past, send...
3    [Following, the, death, of, District, Attorney...
4    [John, Carter, is, a, war-weary,, former, mili...
...
4804 [El, Mariachi, just, wants, to, play, his, gui...
4805 [A, newlywed, couple's, honeymoon, is, upended...
4806 ["Signed,, Sealed,, Delivered", introduces, a,...
4807 [When, ambitious, New, York, attorney, Sam, is...
4808 [Ever, since, the, second, grade, when, he, fi...
Name: overview, Length: 4806, dtype: object
```


In [42]:

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

In [43]:

```
movies.head()
```

Out[43]:

	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 [44]:

```
movies['tags']=movies['overview']+movies['genres']+movies['keywords']+movies['cast']+movies['crew']
```

In [45]:

```
movies.head()
```

Out[45]:

	movie_id	title	overview	genres	keywords	cast	crew	tags
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]	[In, the, 22nd, century,, a, paraplegic, Marin...
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]	[Captain, Barbossa,, long, believed, to, be, d...
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]	[A, cryptic, message, from, Bond's, past, send...
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]	[Following, the, death, of, District, Attorney...
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]	[John, Carter, is, a, war-weary,, former, mili...

In [46]:

```
new_df=movies[['movie_id','title','tags']]
```

In [47]:

new_df

Out[47]:

	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 [48]:

new_df['tags'].apply(lambda x: " ".join(x))

Out[48]:

```

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 [49]:

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

C:\Users\zoaah\AppData\Local\Temp\ipykernel_3996\487797088.py:1: SettingWithCopyWarning:

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/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

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

In [50]:

```
new_df.head()
```

Out[50]:

	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 [52]:

```
new_df['tags'][0]
```

Out[52]:

'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization. Action Adventure Fantasy Science Fiction culture clash future space war space colony society space travel futuristic romance space alien tribe alien planet cgi marine soldier battle love affair anti war power relations mind and soul 3d Sam Worthington Zoe Saldana Sigourney Weaver James Cameron'

In []: