

```
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_title | overview | popularity | production_con |
|---|-----------|---|-----------------------------|------|---|-------------------|----------------|---|------------|---------------------------|
| 0 | 237000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 18, "name": "Drama"}, {"id": 10751, "name": "Science Fiction"}] | http://www.avatarmovie.com/ | 1995 | [{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "marine"}, {"id": 1465, "name": "paraplegic"}] | en | Avatar | In the 22nd century, a paraplegic Marine is di... | 150.437577 | [{"name": "In Film Partne |

```
In [4]: credits.head()
```

```
Out[4]:
```

| | movie_id | title | cast | crew |
|---|----------|--|--|--|
| 0 | 1995 | Avatar | [{"cast_id": 242, "character": "Jake Sully", "credit_id": "52fe48009251416c750aca23", "de...} | [{"credit_id": "52fe48009251416c750aca23", "de...} |
| 1 | 285 | Pirates of the Caribbean: At World's End | [{"cast_id": 4, "character": "Captain Jack Spa...", "credit_id": "52fe4232c3a36847f800b579", "de...} | [{"credit_id": "52fe4232c3a36847f800b579", "de...} |
| 2 | 206647 | Spectre | [{"cast_id": 1, "character": "James Bond", "credit_id": "54805967c3a36829b5002c41", "de...} | [{"credit_id": "54805967c3a36829b5002c41", "de...} |
| 3 | 49026 | The Dark Knight Rises | [{"cast_id": 2, "character": "Bruce Wayne / Ba...", "credit_id": "52fe4781c3a36847f81398c3", "de...} | [{"credit_id": "52fe4781c3a36847f81398c3", "de...} |
| 4 | 49529 | John Carter | [{"cast_id": 5, "character": "John Carter", "credit_id": "52fe479ac3a36847f813eaa3", "de...} | [{"credit_id": "52fe479ac3a36847f813eaa3", "de...} |

marge both data set

```
In [5]: movies.merge(credits, on='title')
```

Out[5]:

| | budget | genres | homepage | id | keywords | original_language | original_title |
|------|-----------|--|--|--------|--|-------------------|--|
| 0 | 237000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}] | http://www.avatarmovie.com/ | 19995 | [{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}] | en | Avatar |
| 1 | 300000000 | [{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}] | http://disney.go.com/disneypictures/pirates/ | 285 | [{"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}] | en | Pirates of the Caribbean: At World's End |
| 2 | 245000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}] | http://www.sonypictures.com/movies/spectre/ | 206647 | [{"id": 470, "name": "spy"}, {"id": 818, "name": "thriller"}] | en | Spectre |
| 3 | 250000000 | [{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}] | http://www.thedarkknightises.com/ | 49026 | [{"id": 849, "name": "dc comics"}, {"id": 853, "name": "superhero"}] | en | The Dark Knight Rises |
| 4 | 260000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}] | http://movies.disney.com/john-carter | 49529 | [{"id": 818, "name": "based on novel"}, {"id": 1464, "name": "culture clash"}] | en | John Carter |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 4804 | 220000 | [{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}] | NaN | 9367 | [{"id": 5616, "name": "united states\u2013mexico"}] | es | El Mariachi |
| 4805 | 9000 | [{"id": 35, "name": "Comedy"}] | NaN | 72766 | [] | en | Newlyweds |

| | budget | genres | homepage | id | keywords | original_language | original_title |
|------|--------|---|---|--------|--|-------------------|---------------------------------|
| | | "Comedy"}, {"id": 10749, "... | | | | | |
| 4806 | 0 | [{"id": 35, "name": "Comedy"}, {"id": 18, "nam... | http://www.hallmarkchannel.com/signedsealledel... | 231617 | [{"id": 248, "name": "date"}, {"id": 699, "nam... | en | Signed, Sealed, Delivered |
| 4807 | 0 | [] | http://shanghaicalling.com/ | 126186 | [] | en | Shanghai Calling |
| 4808 | 0 | [{"id": 99, "name": "Documentary"}] | NaN | 25975 | [{"id": 1523, "name": "obsession"}, {"id": 224... | en | My Date with Drew |
| 4809 | 0 | [{"id": 10749, "name": "Comedy"}, {"id": 18, "name": "Comedy"}] | http://www.hallmarkchannel.com/signedsealledel... | 231617 | [{"id": 248, "name": "date"}, {"id": 699, "name": "date"}] | en | Signed, Sealed, Delivered |

```
In [6]: movies.merge(credits, on='title').shape
```

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

```
In [7]: movies.shape
```

```
Out[7]: (4803, 20)
```

```
In [8]: credits.shape
```

```
Out[8]: (4803, 4)
```

```
In [9]: movies = movies.merge(credits, on='title')
```

```
In [10]: movies.head(1)
```



```
Out[11]: en      4510  
         fr       70  
         es       32  
         zh       27  
         de       27  
         hi       19  
         ja       16  
         it       14  
         ko       12  
         cn       12  
         ru       11  
         pt        9  
         da        7  
         sv        5  
         nl        4  
         fa        4  
         th        3  
         he        3  
         ta        2  
         cs        2  
         ro        2  
         id        2  
         ar        2  
         vi        1  
         sl        1  
         ps        1  
         no        1  
         ky        1  
         hu        1  
         pl        1  
         af        1  
         nb        1  
         tr        1  
         is        1  
         xx        1  
         te        1  
         el        1  
         Name: original_language, dtype: int64
```

```
In [12]: movies.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 4809 entries, 0 to 4808
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  -
0   budget                4809 non-null   int64
1   genres                4809 non-null   object
2   homepage              1713 non-null   object
3   id                    4809 non-null   int64
4   keywords              4809 non-null   object
5   original_language     4809 non-null   object
6   original_title        4809 non-null   object
7   overview              4806 non-null   object
8   popularity            4809 non-null   float64
9   production_companies  4809 non-null   object
10  production_countries  4809 non-null   object
11  release_date          4808 non-null   object
12  revenue               4809 non-null   int64
13  runtime               4807 non-null   float64
14  spoken_languages      4809 non-null   object
15  status                4809 non-null   object
16  tagline               3965 non-null   object
17  title                 4809 non-null   object
18  vote_average          4809 non-null   float64
19  vote_count            4809 non-null   int64
20  movie_id              4809 non-null   int64
21  cast                  4809 non-null   object
22  crew                  4809 non-null   object
dtypes: float64(3), int64(5), object(15)
memory usage: 901.7+ KB

```

```
In [13]: movies = movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```

```
In [14]: movies.head()
```

Out[14]:

| | 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": "52fe479ac3a36847f813eaa3", "de... |

Preprocessing

In [15]: *#Check the missing value*
`movies.isnull().sum()`

Out[15]:

```

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

```

In [17]: *# Three(3) movies that do not have overview so we drop them*
`movies.dropna(inplace=True)`

```
In [18]: #check the duplicate data  
movies.duplicated().sum()
```

```
Out[18]: 0
```

```
In [19]: movies.iloc[0].genres
```

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

```
In [20]: #string given error so we have to convert string of list into list so we import ast library  
import ast
```

```
In [21]: def convert(obj):  
        L = []  
        for i in ast.literal_eval(obj):  
            L.append(i['name'])  
        return L
```

```
In [22]: movies['genres'] = movies['genres'].apply(convert)
```

```
In [23]: movies.head()
```


Out[23]:

| | 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 [24]: `movies['keywords'] = movies['keywords'].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] | [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": "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'] = movies['cast'].apply(convert3)
```

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... | [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 [29]:

```
movies['crew'][0]
```

```

Out[29]: '[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": "Editor", "name":
"Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Pr
oduction Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0,
"id": 900, "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "departmen
t": "Sound", "gender": 0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "53
9c4a4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"},
{"credit_id": "5544ee3b925141499f0008fc", "department": "Sound", "gender": 2, "id": 1729, "job": "Original Music Compos
er", "name": "James Horner"}, {"credit_id": "52fe48009251416c750ac9c3", "department": "Directing", "gender": 2, "id": 2
710, "job": "Director", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "g
ender": 2, "id": 2710, "job": "Writer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "departmen
t": "Editing", "gender": 2, "id": 2710, "job": "Editor", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac
a29", "department": "Production", "gender": 2, "id": 2710, "job": "Producer", "name": "James Cameron"}, {"credit_id":
"52fe48009251416c750aca3f", "department": "Writing", "gender": 2, "id": 2710, "job": "Screenplay", "name": "James Camer
on"}, {"credit_id": "539c4987c3a36810ba0021a4", "department": "Art", "gender": 2, "id": 7236, "job": "Art Direction",
"name": "Andrew Menzies"}, {"credit_id": "549598c3c3a3686ae9004383", "department": "Visual Effects", "gender": 0, "id":
6690, "job": "Visual Effects Producer", "name": "Jill Brooks"}, {"credit_id": "52fe48009251416c750aca4b", "department":
"Production", "gender": 1, "id": 6347, "job": "Casting", "name": "Margery Simkin"}, {"credit_id": "570b6f419251417da700
32fe", "department": "Art", "gender": 2, "id": 6878, "job": "Supervising Art Director", "name": "Kevin Ishioka"}, {"cre
dit_id": "5495a0fac3a3686ae9004468", "department": "Sound", "gender": 0, "id": 6883, "job": "Music Editor", "name": "Di
ck Bernstein"}, {"credit_id": "54959706c3a3686af3003e81", "department": "Sound", "gender": 0, "id": 8159, "job": "Sound
Effects Editor", "name": "Shannon Mills"}, {"credit_id": "54491d58c3a3680fb1001ccb", "department": "Sound", "gender":
0, "id": 8160, "job": "Foley", "name": "Dennie Thorpe"}, {"credit_id": "54491d6cc3a3680fa5001b2c", "department": "Soun
d", "gender": 0, "id": 8163, "job": "Foley", "name": "Jana Vance"}, {"credit_id": "52fe48009251416c750aca57", "departme
nt": "Costume & Make-Up", "gender": 1, "id": 8527, "job": "Costume Design", "name": "Deborah Lynn Scott"}, {"credit_i
d": "52fe48009251416c750aca2f", "department": "Production", "gender": 2, "id": 8529, "job": "Producer", "name": "Jon La
ndau"}, {"credit_id": "539c4937c3a36810ba002194", "department": "Art", "gender": 0, "id": 9618, "job": "Art Direction",
"name": "Sean Haworth"}, {"credit_id": "539c49b6c3a36810c10020e6", "department": "Art", "gender": 1, "id": 12653, "jo
b": "Set Decoration", "name": "Kim Sinclair"}, {"credit_id": "570b6f2f9251413a0e00020d", "department": "Art", "gender":
1, "id": 12653, "job": "Supervising Art Director", "name": "Kim Sinclair"}, {"credit_id": "54491a6c0e0a26748c001b19",
"department": "Art", "gender": 2, "id": 14350, "job": "Set Designer", "name": "Richard F. Mays"}, {"credit_id": "56928c
f4c3a3684cff0025c4", "department": "Production", "gender": 1, "id": 20294, "job": "Executive Producer", "name": "Laeta
Kalogridis"}, {"credit_id": "52fe48009251416c750aca51", "department": "Costume & Make-Up", "gender": 0, "id": 17675, "j
ob": "Costume Design", "name": "Mayes C. Rubeo"}, {"credit_id": "52fe48009251416c750aca11", "department": "Camera", "ge
nder": 2, "id": 18265, "job": "Director of Photography", "name": "Mauro Fiore"}, {"credit_id": "5449194d0e0a26748f001b3
9", "department": "Art", "gender": 0, "id": 42281, "job": "Set Designer", "name": "Scott Herbertson"}, {"credit_id": "5
2fe48009251416c750aca05", "department": "Crew", "gender": 0, "id": 42288, "job": "Stunts", "name": "Woody Schultz"},
{"credit_id": "5592aefb92514152de0010f5", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Makeup A
rtist", "name": "Linda DeVetta"}, {"credit_id": "5592afa492514152de00112c", "department": "Costume & Make-Up", "gende
r": 0, "id": 29067, "job": "Hairstylist", "name": "Linda DeVetta"}, {"credit_id": "54959ed592514130fc002e5d", "departme
nt": "Camera", "gender": 2, "id": 33302, "job": "Camera Operator", "name": "Richard Bluck"}, {"credit_id": "539c4891c3a
36810ba002147", "department": "Art", "gender": 2, "id": 33303, "job": "Art Direction", "name": "Simon Bright"}, {"credi
t_id": "54959c069251417a81001f3a", "department": "Visual Effects", "gender": 0, "id": 113145, "job": "Visual Effects Su
pervisor", "name": "Richard Martin"}, {"credit_id": "54959a0dc3a3680ff5002c8d", "department": "Crew", "gender": 2, "i
d": 58188, "job": "Visual Effects Editor", "name": "Steve R. Moore"}, {"credit_id": "52fe48009251416c750aca1d", "depart
ment": "Editing", "gender": 2, "id": 58871, "job": "Editor", "name": "John Refoua"}, {"credit_id": "54491a4dc3a3680fc30

```

018ca", "department": "Art", "gender": 0, "id": 92359, "job": "Set Designer", "name": "Karl J. Martin"}, {"credit_id": "52fe48009251416c750aca35", "department": "Camera", "gender": 1, "id": 72201, "job": "Director of Photography", "name": "Chiling Lin"}, {"credit_id": "52fe48009251416c750ac9ff", "department": "Crew", "gender": 0, "id": 89714, "job": "Stunt s", "name": "Ilram Choi"}, {"credit_id": "54959c529251416e2b004394", "department": "Visual Effects", "gender": 2, "id": 93214, "job": "Visual Effects Supervisor", "name": "Steven Quale"}, {"credit_id": "54491edf0e0a267489001c37", "department": "Crew", "gender": 1, "id": 122607, "job": "Dialect Coach", "name": "Carla Meyer"}, {"credit_id": "539c485bc3a368653d001a3a", "department": "Art", "gender": 2, "id": 132585, "job": "Art Direction", "name": "Nick Bassett"}, {"credit_id": "539c4903c3a368653d001a74", "department": "Art", "gender": 0, "id": 132596, "job": "Art Direction", "name": "Jill Cormack"}, {"credit_id": "539c4967c3a368653d001a94", "department": "Art", "gender": 0, "id": 132604, "job": "Art Direction", "name": "Andy McLaren"}, {"credit_id": "52fe48009251416c750aca45", "department": "Crew", "gender": 0, "id": 236696, "job": "Motion Capture Artist", "name": "Terry Notary"}, {"credit_id": "54959e02c3a3680fc60027d2", "department": "Crew", "gender": 2, "id": 956198, "job": "Stunt Coordinator", "name": "Garrett Warren"}, {"credit_id": "54959ca3c3a3686ae300438c", "department": "Visual Effects", "gender": 2, "id": 957874, "job": "Visual Effects Supervisor", "name": "Jonathan Rothbart"}, {"credit_id": "570b6f519251412c74001b2f", "department": "Art", "gender": 0, "id": 957889, "job": "Supervising Art Director", "name": "Stefan Dechant"}, {"credit_id": "570b6f62c3a3680b77007460", "department": "Art", "gender": 2, "id": 959555, "job": "Supervising Art Director", "name": "Todd Cherniawsky"}, {"credit_id": "539c4a3ac3a36810da0021cc", "department": "Production", "gender": 0, "id": 1016177, "job": "Casting", "name": "Miranda Rivers"}, {"credit_id": "539c482cc3a36810c1002062", "department": "Art", "gender": 0, "id": 1032536, "job": "Production Design", "name": "Robert Stromberg"}, {"credit_id": "539c4b65c3a36810c9002125", "department": "Costume & Make-Up", "gender": 2, "id": 1071680, "job": "Costume Design", "name": "John Harding"}, {"credit_id": "54959e6692514130fc002e4e", "department": "Camera", "gender": 0, "id": 1177364, "job": "Steadicam Operator", "name": "Roberto De Angelis"}, {"credit_id": "539c49f1c3a368653d001aac", "department": "Costume & Make-Up", "gender": 2, "id": 1202850, "job": "Makeup Department Head", "name": "Mike Smithson"}, {"credit_id": "5495999ec3a3686ae100460c", "department": "Visual Effects", "gender": 0, "id": 1204668, "job": "Visual Effects Producer", "name": "Alain Lalanne"}, {"credit_id": "54959cdfc3a3681153002729", "department": "Visual Effects", "gender": 0, "id": 1206410, "job": "Visual Effects Supervisor", "name": "Lucas Salton"}, {"credit_id": "549596239251417a81001eae", "department": "Crew", "gender": 0, "id": 1234266, "job": "Post Production Supervisor", "name": "Janace Tashjian"}, {"credit_id": "54959c859251416e1e003efe", "department": "Visual Effects", "gender": 0, "id": 1271932, "job": "Visual Effects Supervisor", "name": "Stephen Rosenbaum"}, {"credit_id": "5592af28c3a368775a00105f", "department": "Costume & Make-Up", "gender": 0, "id": 1310064, "job": "Makeup Artist", "name": "Frankie Karena"}, {"credit_id": "539c4adfc3a36810e300203b", "department": "Costume & Make-Up", "gender": 1, "id": 1319844, "job": "Costume Supervisor", "name": "Lisa Lovaas"}, {"credit_id": "54959b579251416e2b004371", "department": "Visual Effects", "gender": 0, "id": 1327028, "job": "Visual Effects Supervisor", "name": "Jonathan Fawcner"}, {"credit_id": "539c48a7c3a36810b5001fa7", "department": "Art", "gender": 0, "id": 1330561, "job": "Art Direction", "name": "Robert Bavin"}, {"credit_id": "539c4a71c3a36810da0021e0", "department": "Costume & Make-Up", "gender": 0, "id": 1330567, "job": "Costume Supervisor", "name": "Anthony Almaraz"}, {"credit_id": "539c4a8ac3a36810ba0021e4", "department": "Costume & Make-Up", "gender": 0, "id": 1330570, "job": "Costume Supervisor", "name": "Carolyn M. Fenton"}, {"credit_id": "539c4ab6c3a36810da0021f0", "department": "Costume & Make-Up", "gender": 0, "id": 1330574, "job": "Costume Supervisor", "name": "Beth Koenigsberg"}, {"credit_id": "54491ab70e0a267480001ba2", "department": "Art", "gender": 0, "id": 1336191, "job": "Set Designer", "name": "Sam Page"}, {"credit_id": "544919d9c3a3680fc30018bd", "department": "Art", "gender": 0, "id": 1339441, "job": "Set Designer", "name": "Tex Kadonaga"}, {"credit_id": "54491cf50e0a267483001b0c", "department": "Editing", "gender": 0, "id": 1352422, "job": "Dialogue Editor", "name": "Kim Foscatto"}, {"credit_id": "544919f40e0a26748c001b09", "department": "Art", "gender": 0, "id": 1352962, "job": "Set Designer", "name": "Tammy S. Lee"}, {"credit_id": "5495a115c3a3680ff5002d71", "department": "Crew", "gender": 0, "id": 1357070, "job": "Transportation Coordinator", "name": "Denny Cairra"}, {"credit_id": "5495a12f92514130fc002e94", "department": "Crew", "gender": 0, "id": 1357071, "job": "Transportation Coordinator", "name": "James Waitkus"}, {"credit_id": "5495976fc3a36811530026b0", "department": "Sound", "gender": 0, "id": 13601

03, "job": "Supervising Sound Editor", "name": "Addison Teague"}, {"credit_id": "54491837c3a3680fb1001c5a", "department": "Art", "gender": 2, "id": 1376887, "job": "Set Designer", "name": "C. Scott Baker"}, {"credit_id": "54491878c3a3680fb4001c9d", "department": "Art", "gender": 0, "id": 1376888, "job": "Set Designer", "name": "Luke Caska"}, {"credit_id": "544918dac3a3680fa5001ae0", "department": "Art", "gender": 0, "id": 1376889, "job": "Set Designer", "name": "David Chow"}, {"credit_id": "544919110e0a267486001b68", "department": "Art", "gender": 0, "id": 1376890, "job": "Set Designer", "name": "Jonathan Dyer"}, {"credit_id": "54491967c3a3680faa001b5e", "department": "Art", "gender": 0, "id": 1376891, "job": "Set Designer", "name": "Joseph Hiura"}, {"credit_id": "54491997c3a3680fb1001c8a", "department": "Art", "gender": 0, "id": 1376892, "job": "Art Department Coordinator", "name": "Rebecca Jellie"}, {"credit_id": "544919ba0e0a26748f001b42", "department": "Art", "gender": 0, "id": 1376893, "job": "Set Designer", "name": "Robert Andrew Johnson"}, {"credit_id": "54491b1dc3a3680faa001b8c", "department": "Art", "gender": 0, "id": 1376895, "job": "Assistant Art Director", "name": "Mike Stassi"}, {"credit_id": "54491b79c3a3680fbb001826", "department": "Art", "gender": 0, "id": 1376897, "job": "Construction Coordinator", "name": "John Villarino"}, {"credit_id": "54491baec3a3680fb4001ce6", "department": "Art", "gender": 2, "id": 1376898, "job": "Assistant Art Director", "name": "Jeffrey Wisniewski"}, {"credit_id": "54491d2fc3a3680fb4001d07", "department": "Editing", "gender": 0, "id": 1376899, "job": "Dialogue Editor", "name": "Cheryl Nard"}, {"credit_id": "54491d86c3a3680fa5001b2f", "department": "Editing", "gender": 0, "id": 1376901, "job": "Dialogue Editor", "name": "Marshall Winn"}, {"credit_id": "54491d9dc3a3680faa001bb0", "department": "Sound", "gender": 0, "id": 1376902, "job": "Supervising Sound Editor", "name": "Gwendolyn Yates Whittle"}, {"credit_id": "54491dc10e0a267486001bce", "department": "Sound", "gender": 0, "id": 1376903, "job": "Sound Re-Recording Mixer", "name": "William Stein"}, {"credit_id": "54491f500e0a26747c001c07", "department": "Crew", "gender": 0, "id": 1376909, "job": "Choreographer", "name": "Lula Washington"}, {"credit_id": "549599239251412c4e002a2e", "department": "Visual Effects", "gender": 0, "id": 1391692, "job": "Visual Effects Producer", "name": "Chris Del Conte"}, {"credit_id": "54959d54c3a36831b8001d9a", "department": "Visual Effects", "gender": 2, "id": 1391695, "job": "Visual Effects Supervisor", "name": "R. Christopher White"}, {"credit_id": "54959bdf9251412c4e002a66", "department": "Visual Effects", "gender": 0, "id": 1394070, "job": "Visual Effects Supervisor", "name": "Dan Lemmon"}, {"credit_id": "5495971d92514132ed002922", "department": "Sound", "gender": 0, "id": 1394129, "job": "Sound Effects Editor", "name": "Tim Nielsen"}, {"credit_id": "5592b25792514152cc0011aa", "department": "Crew", "gender": 0, "id": 1394286, "job": "CG Supervisor", "name": "Michael Mulholland"}, {"credit_id": "54959a329251416e2b004355", "department": "Crew", "gender": 0, "id": 1394750, "job": "Visual Effects Editor", "name": "Thomas Nittmann"}, {"credit_id": "54959d6dc3a3686ae9004401", "department": "Visual Effects", "gender": 0, "id": 1394755, "job": "Visual Effects Supervisor", "name": "Edson Williams"}, {"credit_id": "5495a08fc3a3686ae300441c", "department": "Editing", "gender": 0, "id": 1394953, "job": "Digital Intermediate", "name": "Christine Carr"}, {"credit_id": "55402d659251413d6d000249", "department": "Visual Effects", "gender": 0, "id": 1395269, "job": "Visual Effects Supervisor", "name": "John Bruno"}, {"credit_id": "54959e7b9251416e1e003f3e", "department": "Camera", "gender": 0, "id": 1398970, "job": "Steadicam Operator", "name": "David Emmerichs"}, {"credit_id": "54959734c3a3686ae10045e0", "department": "Sound", "gender": 0, "id": 1400906, "job": "Sound Effects Editor", "name": "Christopher Scarabosio"}, {"credit_id": "549595dd92514130fc002d79", "department": "Production", "gender": 0, "id": 1401784, "job": "Production Supervisor", "name": "Jennifer Teves"}, {"credit_id": "549596009251413af70028cc", "department": "Production", "gender": 0, "id": 1401785, "job": "Production Manager", "name": "Brigitte Yorke"}, {"credit_id": "549596e892514130fc002d99", "department": "Sound", "gender": 0, "id": 1401786, "job": "Sound Effects Editor", "name": "Ken Fischer"}, {"credit_id": "549598229251412c4e002a1c", "department": "Crew", "gender": 0, "id": 1401787, "job": "Special Effects Coordinator", "name": "Iain Hutton"}, {"credit_id": "549598349251416e2b00432b", "department": "Crew", "gender": 0, "id": 1401788, "job": "Special Effects Coordinator", "name": "Steve Ingram"}, {"credit_id": "54959905c3a3686ae3004324", "department": "Visual Effects", "gender": 0, "id": 1401789, "job": "Visual Effects Producer", "name": "Joyce Cox"}, {"credit_id": "5495994b92514132ed002951", "department": "Visual Effects", "gender": 0, "id": 1401790, "job": "Visual Effects Producer", "name": "Jenny Foster"}, {"credit_id": "549599cb3a3686ae1004613", "department": "Crew", "gender": 0, "id": 1401791, "job": "Visual Effects Editor", "name": "Christopher Marino"}, {"credit_id": "549599f2c3a3686ae100461e", "department": "Crew", "gender": 0, "id": 1401792,

```

"job": "Visual Effects Editor", "name": "Jim Milton"}, {"credit_id": "54959a51c3a3686af3003eb5", "department": "Visual Effects", "gender": 0, "id": 1401793, "job": "Visual Effects Producer", "name": "Cyndi Ochs"}, {"credit_id": "54959a7cc3a36811530026f4", "department": "Crew", "gender": 0, "id": 1401794, "job": "Visual Effects Editor", "name": "Lucas Putnam"}, {"credit_id": "54959b91c3a3680ff5002cb4", "department": "Visual Effects", "gender": 0, "id": 1401795, "job": "Visual Effects Supervisor", "name": "Anthony 'Max' Ivins"}, {"credit_id": "54959bb69251412c4e002a5f", "department": "Visual Effects", "gender": 0, "id": 1401796, "job": "Visual Effects Supervisor", "name": "John Knoll"}, {"credit_id": "54959cbbc3a3686ae3004391", "department": "Visual Effects", "gender": 2, "id": 1401799, "job": "Visual Effects Supervisor", "name": "Eric Saindon"}, {"credit_id": "54959d06c3a3686ae90043f6", "department": "Visual Effects", "gender": 0, "id": 1401800, "job": "Visual Effects Supervisor", "name": "Wayne Stables"}, {"credit_id": "54959d259251416e1e003f11", "department": "Visual Effects", "gender": 0, "id": 1401801, "job": "Visual Effects Supervisor", "name": "David Stinnett"}, {"credit_id": "54959db49251413af7002975", "department": "Visual Effects", "gender": 0, "id": 1401803, "job": "Visual Effects Supervisor", "name": "Guy Williams"}, {"credit_id": "54959de4c3a3681153002750", "department": "Crew", "gender": 0, "id": 1401804, "job": "Stunt Coordinator", "name": "Stuart Thorp"}, {"credit_id": "54959ef2c3a3680fc60027f2", "department": "Lighting", "gender": 0, "id": 1401805, "job": "Best Boy Electric", "name": "Giles Coburn"}, {"credit_id": "54959f07c3a3680fc60027f9", "department": "Camera", "gender": 2, "id": 1401806, "job": "Still Photographer", "name": "Mark Fellman"}, {"credit_id": "54959f47c3a3681153002774", "department": "Lighting", "gender": 0, "id": 1401807, "job": "Lighting Technician", "name": "Scott Sprague"}, {"credit_id": "54959f8cc3a36831b8001df2", "department": "Visual Effects", "gender": 0, "id": 1401808, "job": "Animation Director", "name": "Jeremy Hollobon"}, {"credit_id": "54959fa0c3a36831b8001dfb", "department": "Visual Effects", "gender": 0, "id": 1401809, "job": "Animation Director", "name": "Orlando Meunier"}, {"credit_id": "54959fb6c3a3686af3003f54", "department": "Visual Effects", "gender": 0, "id": 1401810, "job": "Animation Director", "name": "Taisuke Tanimura"}, {"credit_id": "54959fd2c3a36831b8001e02", "department": "Costume & Make-Up", "gender": 0, "id": 1401812, "job": "Set Costumer", "name": "Lilia Mishel Acevedo"}, {"credit_id": "54959ff9c3a3686ae300440c", "department": "Costume & Make-Up", "gender": 0, "id": 1401814, "job": "Set Costumer", "name": "Alejandro M. Hernandez"}, {"credit_id": "5495a0ddc3a3686ae10046fe", "department": "Editing", "gender": 0, "id": 1401815, "job": "Digital Intermediate", "name": "Marvin Hall"}, {"credit_id": "5495a1f7c3a3686ae3004443", "department": "Production", "gender": 0, "id": 1401816, "job": "Publicist", "name": "Judy Alley"}, {"credit_id": "5592b29fc3a36869d100002f", "department": "Crew", "gender": 0, "id": 1418381, "job": "CG Supervisor", "name": "Mike Perry"}, {"credit_id": "5592b23a9251415df8001081", "department": "Crew", "gender": 0, "id": 1426854, "job": "CG Supervisor", "name": "Andrew Morley"}, {"credit_id": "55491e1192514104c40002d8", "department": "Art", "gender": 0, "id": 1438901, "job": "Conceptual Design", "name": "Seth Engstrom"}, {"credit_id": "5525d5809251417276002b06", "department": "Crew", "gender": 0, "id": 1447362, "job": "Visual Effects Art Director", "name": "Eric Oliver"}, {"credit_id": "554427ca925141586500312a", "department": "Visual Effects", "gender": 0, "id": 1447503, "job": "Modeling", "name": "Matsune Suzuki"}, {"credit_id": "551906889251415aab001c88", "department": "Art", "gender": 0, "id": 1447524, "job": "Art Department Manager", "name": "Paul Tobin"}, {"credit_id": "5592af8492514152cc0010de", "department": "Costume & Make-Up", "gender": 0, "id": 1452643, "job": "Hairstylist", "name": "Roxane Griffin"}, {"credit_id": "553d3c109251415852001318", "department": "Lighting", "gender": 0, "id": 1453938, "job": "Lighting Artist", "name": "Arun Ram-Mohan"}, {"credit_id": "5592af4692514152d5001355", "department": "Costume & Make-Up", "gender": 0, "id": 1457305, "job": "Makeup Artist", "name": "Georgia Lockhart-Adams"}, {"credit_id": "5592b2eac3a36877470012a5", "department": "Crew", "gender": 0, "id": 1466035, "job": "CG Supervisor", "name": "Thrain Shadbolt"}, {"credit_id": "5592b032c3a36877450015f1", "department": "Crew", "gender": 0, "id": 1483220, "job": "CG Supervisor", "name": "Brad Alexander"}, {"credit_id": "5592b05592514152d80012f6", "department": "Crew", "gender": 0, "id": 1483221, "job": "CG Supervisor", "name": "Shadi Almassizadeh"}, {"credit_id": "5592b090c3a36877570010b5", "department": "Crew", "gender": 0, "id": 1483222, "job": "CG Supervisor", "name": "Simon Clutterbuck"}, {"credit_id": "5592b0dbc3a368774b00112c", "department": "Crew", "gender": 0, "id": 1483223, "job": "CG Supervisor", "name": "Graeme Demmocks"}, {"credit_id": "5592b0fe92514152db0010c1", "department": "Crew", "gender": 0, "id": 1483224, "job": "CG Supervisor", "name": "Adrian Fernandes"}, {"credit_id": "5592b11f9251415df8001059", "department": "Crew", "gender": 0, "id": 1483225, "job":

```

```
"CG Supervisor", "name": "Mitch Gates"}, {"credit_id": "5592b15dc3a3687745001645", "department": "Crew", "gender": 0,
"id": 1483226, "job": "CG Supervisor", "name": "Jerry Kung"}, {"credit_id": "5592b18e925141645a0004ae", "department":
"Crew", "gender": 0, "id": 1483227, "job": "CG Supervisor", "name": "Andy Lomas"}, {"credit_id": "5592b1bfc3a368775d001
0e7", "department": "Crew", "gender": 0, "id": 1483228, "job": "CG Supervisor", "name": "Sebastian Marino"}, {"credit_i
d": "5592b2049251415df8001078", "department": "Crew", "gender": 0, "id": 1483229, "job": "CG Supervisor", "name": "Matt
hias Menz"}, {"credit_id": "5592b27b92514152d800136a", "department": "Crew", "gender": 0, "id": 1483230, "job": "CG Sup
ervisor", "name": "Sergei Nevshupov"}, {"credit_id": "5592b2c3c3a36869e800003c", "department": "Crew", "gender": 0, "i
d": 1483231, "job": "CG Supervisor", "name": "Philippe Rebours"}, {"credit_id": "5592b317c3a36877470012af", "departmen
t": "Crew", "gender": 0, "id": 1483232, "job": "CG Supervisor", "name": "Michael Takarangi"}, {"credit_id": "5592b345c3
a36877470012bb", "department": "Crew", "gender": 0, "id": 1483233, "job": "CG Supervisor", "name": "David Weitzberg"},
{"credit_id": "5592b37cc3a368775100113b", "department": "Crew", "gender": 0, "id": 1483234, "job": "CG Supervisor", "na
me": "Ben White"}, {"credit_id": "573c8e2f9251413f5d000094", "department": "Crew", "gender": 1, "id": 1621932, "job":
"Stunts", "name": "Min Windle"}]'
```

In [30]: *#dataset column data is in dictionary format so we want to extract job key from crew column*

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

In [31]: `movies['crew'] = movies['crew'].apply(fetch_director)`

In [32]: `movies.head()`

Out[32]:

| | 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 [33]: `movies['overview'][0]`

Out[33]: '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 [34]: `#overview column is in string so we convert it into list`
`movies['overview'].apply(lambda x: x.split())`

Out[34]:

```

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

Out[35]:

| | 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] |

Apply Transformation

In [36]: `'''remove the space between the words like sam worthington so this one full name but it count as a seperate tag so we have to remove the sapce between the full name like samworthington'''`

```
movies['genres'] = movies['genres'].apply(lambda x:[i.replace(" ", "") for i in x])
movies['keywords'] = movies['keywords'].apply(lambda x:[i.replace(" ", "") for i in x])
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 [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, ScienceFiction] | [cultureclash, future, spacewar, spacecolony, ... | [SamWorthington, ZoeSaldana, SigourneyWeaver] | [JamesCameron] |
| 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] | [GoreVerbinski] |
| 2 | 206647 | Spectre | [A, cryptic, message, from, Bond's, past, send... | [Action, Adventure, Crime] | [spy, basedonnovel, secretagent, sequel, mi6, ... | [DanielCraig, ChristophWaltz, LéaSeydoux] | [SamMendes] |
| 3 | 49026 | The Dark Knight Rises | [Following, the, death, of, District, Attorney... | [Action, Crime, Drama, Thriller] | [dcomics, crimefighter, terrorist, secretiden... | [ChristianBale, MichaelCaine, GaryOldman] | [ChristopherNolan] |
| 4 | 49529 | John Carter | [John, Carter, is, a, war-weary,, former, mili... | [Action, Adventure, ScienceFiction] | [basedonnovel, mars, medallion, spacetravel, p... | [TaylorKitsch, LynnCollins, SamanthaMorton] | [AndrewStanton] |

In [50]:

```
#create a new column and concatenate the all column in it
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
```

In [51]:

```
movies.head()
```

Out[51]:

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

```
In [52]: # crete a new data frame with movie id ,title and tags
new_df = movies[['movie_id', 'title', 'tags']]
```

```
In [56]: new_df['tags'] = new_df['tags'].apply(lambda x:" ".join(x))
```

C:\Users\lenovo\AppData\Local\Temp\ipykernel_8244\3089450492.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
new_df['tags'] = new_df['tags'].apply(lambda x:" ".join(x))

```
In [57]: new_df.head()
```

Out[57]:

| | 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 [71]: `#import after the vectors_features_names`
`import nltk`

In [72]: `from nltk.stem.porter import PorterStemmer`
`ps = PorterStemmer()`

In [77]: `def stem(text):`
 `y = []`
 `for i in text.split():`
 `y.append(ps.stem(i))`
 `return " ".join(y)`

In [79]: `new_df['tags'] = new_df['tags'].apply(stem)`

C:\Users\lenovo\AppData\Local\Temp\ipykernel_8244\3213734980.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
`new_df['tags'] = new_df['tags'].apply(stem)`

In [58]: `new_df['tags'][0]`

Out[58]: '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 ScienceFiction cultureclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle love affair antiwar powerrelations mindandsoul 3d SamWorthington Zoesaldana SigourneyWeaver JamesCameron'

In [59]: `new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())`

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_8244\3214958533.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
new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

```
In [60]: new_df.head()
```

```
Out[60]:
```

| | 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 [61]: new_df['tags'][0]
```

```
Out[61]: 'in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but becomes torn betwe
en following orders and protecting an alien civilization. action adventure fantasy sciencefiction cultureclash future s
pacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle love
affair antiwar powerrelations mindandsoul 3d samworthington zoesaldana sigourneyweaver jamescameron'
```

```
In [62]: new_df['tags'][1]
```

```
Out[62]: "captain barbossa, long believed to be dead, has come back to life and is headed to the edge of the earth with will tur
ner and elizabeth swann. but nothing is quite as it seems. adventure fantasy action ocean drugabuse exoticisland eastin
diatradingcompany loveofone'slife traitor shipwreck strongwoman ship alliance calypso afterlife fighter pirate swashbuc
kler aftercreditsstinger johnnydepp orlandobloom keiraknightley goreverbinski"
```

Feature Extraction

```
In [80]: from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000, stop_words='english')
```

```
In [81]: # cv.fit_transform(new_df['tags']).toarray()
#cv.fit_transform(new_df['tags']).toarray().shape
```

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

```
In [82]: vectors
```

```
Out[82]: 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 [83]: vectors[0]
```

```
Out[83]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
In [84]: cv.get_feature_names()
```

```
C:\Users\lenovo\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names
is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out inst
ead.
  warnings.warn(msg, category=FutureWarning)
```

```
Out[84]: ['000',  
          '007',  
          '10',  
          '100',  
          '11',  
          '12',  
          '13',  
          '14',  
          '15',  
          '16',  
          '17',  
          '17th',  
          '18',  
          '18th',  
          '18thcenturi',  
          '19',  
          '1910',  
          '1920',  
          '1930',  
          '1940',  
          '1944',  
          '1950',  
          '1950s',  
          '1960',  
          '1960s',  
          '1970',  
          '1970s',  
          '1971',  
          '1974',  
          '1976',  
          '1980',  
          '1985',  
          '1990',  
          '1999',  
          '19th',  
          '19thcenturi',  
          '20',  
          '200',  
          '2003',  
          '2009',  
          '20th',  
          '21st',  
          '23',  
          '24',  
          '25',
```


'30',
'300',
'3d',
'40',
'50',
'500',
'60',
'70',
'80',
'aaron',
'aaroneckhart',
'abandon',
'abduct',
'abigailbreslin',
'abil',
'abl',
'aboard',
'abov',
'abus',
'academ',
'academi',
'accept',
'access',
'accid',
'accident',
'acclaim',
'accompani',
'accomplish',
'account',
'accus',
'ace',
'achiev',
'acquaint',
'act',
'action',
'actionhero',
'activ',
'activist',
'activities',
'actor',
'actress',
'actual',
'ad',
'adam',
'adamsandl',

'adamshankman',
'adapt',
'add',
'addict',
'adjust',
'admir',
'admit',
'adolesc',
'adopt',
'ador',
'adrienbrodi',
'adult',
'adultanim',
'adulteri',
'adulthood',
'advanc',
'adventur',
'adventure',
'adventures',
'advertis',
'advic',
'advis',
'affair',
'affect',
'afghanistan',
'africa',
'african',
'africanamerican',
'aftercreditssting',
'afterlif',
'aftermath',
'ag',
'age',
'agediffer',
'agenc',
'agency',
'agenda',
'agent',
'agents',
'aggress',
'ago',
'agre',
'ahead',
'aid',
'aidanquinn',

'ail',
'aim',
'air',
'airplan',
'airplanecrash',
'airport',
'aka',
'al',
'alabama',
'alan',
'alaska',
'albert',
'alcatraz',
'alcohol',
'alecbaldwin',
'alex',
'alexxkendrick',
'alfredhitchcock',
'alfredmolina',
'ali',
'alic',
'alice',
'alien',
'alieninvas',
'alienlife',
'alienplanet',
'aliens',
'alik',
'aliv',
'alive',
'allen',
'alli',
'allianc',
'allow',
'alon',
'alongsid',
'alpacino',
'alpha',
'alreadi',
'alter',
'altern',
'alway',
'alyssa',
'alzheimer',
'amanda',

'amandapeet',
'amandaseyfri',
'amateur',
'amaz',
'amazon',
'ambassador',
'ambit',
'ambiti',
'ambul',
'ambush',
'america',
'american',
'americanabroad',
'americancivilwar',
'americanfootball',
'americanfootballplay',
'amid',
'amidst',
'amnesia',
'amp',
'amsterdam',
'amus',
'amusementpark',
'amy',
'amyadam',
'mysmart',
'ana',
'anakin',
'analyst',
'anarchiccomedi',
'ancient',
'ancientrom',
'ancientworld',
'anderson',
'andi',
'andiemacdowel',
'andrew',
'android',
'andy',
'andygarcía',
'angel',
'angela',
'angelabassett',
'angeles',
'angelinajoli',

'anger',
'angle',
'angri',
'ani',
'anim',
'animalattack',
'animalhorror',
'animals',
'anjelicaheston',
'ann',
'anna',
'annafari',
'annakendrick',
'anne',
'annehathaway',
'annemoss',
'annetteben',
'anni',
'annie',
'anniversari',
'announc',
'annual',
'anonym',
'anoth',
'answer',
'ant',
'antholog',
'anthoni',
'anthonyanderson',
'anthonyhopkin',
'anthonymacki',
'anthropomorph',
'anti',
'antic',
'antihero',
'antiqu',
'antoinefuqua',
'antoniobandera',
'antonyelchin',
'anyon',
'anyth',
'apart',
'apartheid',
'apartment',
'ape',

'apocalyps',
'apocalypse',
'apocalypt',
'appar',
'appear',
'appl',
'apple',
'appoint',
'appreci',
'apprentic',
'approach',
'april',
'aquarium',
'arab',
'arch',
'archaeologist',
'archeolog',
'archer',
'architect',
'arctic',
'area',
'aren',
'arena',
'argument',
'aris',
'aristocrat',
'arm',
'armi',
'armor',
'armsdeal',
'army',
'arnold',
'arnoldschwarzenegg',
'arrang',
'arrangedmarriag',
'arrest',
'arriv',
'arrog',
'art',
'arthur',
'artifact',
'artifici',
'artificialintellig',
'artist',
'ash',

'ashley',
'ashleyjudd',
'ashtonkutch',
'asia',
'asian',
'asid',
'ask',
'aspect',
'aspir',
'assassin',
'assault',
'assembl',
'assign',
'assist',
'assistant',
'associ',
'assum',
'asteroid',
'astronaut',
'asylum',
'atheist',
'athlet',
'atom',
'atomicbomb',
'attack',
'attacks',
'attempt',
'attempttoescap',
'attend',
'attent',
'attic',
'attitud',
'attorney',
'attract',
'auction',
'audienc',
'audit',
'august',
'aunt',
'austin',
'australia',
'australian',
'author',
'autism',
'auto',

'automobilerac',
'aveng',
'averag',
'avoid',
'await',
'awak',
'awaken',
'award',
'away',
'awkward',
'awri',
'awry',
'ax',
'babe',
'babi',
'baby',
'bachelor',
'backdrop',
'background',
'backpack',
'bad',
'bag',
'bahama',
'bail',
'balanc',
'ball',
'ballet',
'balloon',
'ban',
'band',
'bandit',
'bangkok',
'banish',
'bank',
'banker',
'bankrobb',
'bankrobber',
'bar',
'barbrastreisand',
'bare',
'bargain',
'barn',
'barney',
'baron',
'barri',

'barrylevinson',
'barrysonnenfeld',
'bas',
'base',
'basebal',
'basedoncomicbook',
'basedongraphicnovel',
'basedonnovel',
'basedonplay',
'basedonstagemus',
'basedontrueev',
'basedontruestori',
'basedontvseri',
'basedonvideogam',
'basedonyoungadultnovel',
'basement',
'basketbal',
'bat',
'batman',
'battl',
'battle',
'battlefield',
'bay',
'beach',
'beam',
'bear',
'beard',
'beast',
'beat',
'beauti',
'beautiful',
'beautifulwoman',
'beauty',
'becam',
'becaus',
'becki',
'becom',
'becominganadult',
'bed',
'bedroom',
'beer',
'befor',
'befriend',
'began',
'begin',

'begins',
'behavior',
'belief',
'believ',
'bell',
'bella',
'belong',
'belov',
'ben',
'benaffleck',
'bend',
'beneath',
'benefit',
'benfost',
'beniciodeltoro',
'benjamin',
'benjaminbratt',
'benkingsley',
'bennett',
'benstil',
'bent',
'berlin',
'best',
'bestfriend',
'bestfriendsinlov',
'bet',
'beth',
'betray',
'bettemidl',
'better',
'betti',
'beverli',
'bibl',
'bid',
'big',
'bigger',
'biggest',
'bike',
'biker',
'bikini',
'billhad',
'billi',
'billionair',
'billmurray',
'billnighi',

'billpaxton',
'billpullman',
'billybobthornton',
'billycrudup',
'billycryst',
'biographi',
'biolog',
'bird',
'birth',
'birthday',
'bisexu',
'bishop',
'bit',
'bite',
'bitter',
'bizarr',
'black',
'blackmag',
'blackmail',
'blackpeopl',
'blacksmith',
'blade',
'blame',
'blend',
'blind',
'bliss',
'blizzard',
'block',
'blond',
'blood',
'bloodi',
'bloodsplatt',
'bloodthirsti',
'blow',
'blue',
'board',
'boardingschool',
'boat',
'bob',
'bobbi',
'bobbyfarrelli',
'bobhoskin',
'bodi',
'body',
'bodyguard',

'bold',
'bollywood',
'bomb',
'bond',
'bone',
'book',
'border',
'bore',
'boredom',
'born',
'boss',
'boston',
'botch',
'bound',
'boundari',
'bounti',
'bountyhunt',
'box',
'boxer',
'boy',
'boyfriend',
'boys',
'bradleycoop',
'bradpitt',
'brain',
'brand',
'brandon',
'brave',
'braveri',
'brazil',
'brazilian',
'break',
'breakdown',
'breakup',
'breast',
'breath',
'breed',
'brendanfras',
'brendangleeson',
'brent',
'brettratn',
'brian',
'briandepalma',
'bride',
'bridesmaid',

'bridg',
'brief',
'brien',
'bright',
'brilliant',
'bring',
'brink',
'britain',
'british',
'britishsecretservic',
'brittanymurphi',
'broadcast',
'broadway',
'broke',
'broken',
'broker',
'bronx',
'brook',
'brooklyn',
'broom',
'brothel',
'brother',
'brotherbrotherrelationship',
'brothers',
'brothersisterrelationship',
'brought',
'brown',
'bruce',
'brucegreenwood',
'brucewilli',
'brutal',
'bryansing',
'bu',
'buck',
'bud',
'buddi',
'buddy',
'buddycomedi',
'buddycop',
'budget',
'build',
'building',
'built',
'bullet',
'bulli',

'bumbl',
'bunch',
'bunker',
'bunni',
'burglar',
'buri',
'burn',
'bush',
'busi',
'business',
'businessman',
'bust',
'butcher',
'butler',
'butt',
'button',
'buy',
'buzz',
'cabin',
'caesar',
'cage',
'cairo',
'cal',
'california',
'calvin',
'camcord',
'came',
'camera',
'cameraman',
'camerondiaz',
'camp',
'campaign',
'campbell',
'campu',
'canada',
'canadian',
'cancer',
'candi',
'candid',
'canin',
'cannib',
'canuxploit',
'capabl',
'caper',
'capit',

'capt',
'captain',
'captiv',
'captur',
'capture',
'car',
'caraccid',
'carchas',
'carcrash',
'card',
'care',
'career',
'carefre',
'caretak',
'careymulligan',
'caribbean',
'carjourney',
'carl',
'carlagugino',
'carmen',
'carol',
'carolina',
'carrac',
'carri',
'carrie',
'cartel',
'carter',
'cartoon',
'caryelw',
'case',
'caseyaffleck',
'cash',
'casino',
'cast',
'castl',
'cat',
'cataclysm',
'catastroph',
'catch',
'cateblanchett',
'catherinedeneuve',
'catherinekeen',
'catherinezeta',
'cathol',
'catholic',

'cattl',
'caught',
'caus',
'cavalri',
'cave',
'cavemen',
'celebr',
'celebration',
'cell',
'cellphon',
'cemeteri',
'center',
'centr',
'central',
'centuri',
'centuries',
'century',
'ceo',
'certain',
'chad',
'chain',
'chainsaw',
'challeng',
'chamber',
'champion',
'championship',
'chanc',
'chance',
'chang',
'change',
'changed',
'changes',
'channingtatum',
'chao',
'chaos',
'chaotic',
'chapter',
'charact',
'character',
'characters',
'charg',
'charismat',
'charl',
'charli',
'charlie',

'charliesheen',
'charlizetheron',
'charm',
'chart',
'chase',
'chauffeur',
'chazzpalminteri',
'cheat',
'check',
'cheerlead',
'chef',
'chemic',
'cher',
'chevychas',
'chicago',
'chicken',
'chief',
'child',
'childabus',
'childhero',
'childhood',
'childprodigi',
'children',
'chill',
'chimp',
'china',
'chines',
'chip',
'chipmunk',
'chiwetelejofor',
'chloe',
'chloëgracemoretz',
'chloësevigni',
'chocol',
'choic',
'choice',
'choos',
'chosen',
'chowyun',
'chri',
'chriscolumbu',
'chriscoop',
'chrisevan',
'chrishemsworth',
'chrisklein',

'chrispin',
'chrisrock',
'christ',
'christian',
'christianbal',
'christianslat',
'christin',
'christinaappleg',
'christinaricci',
'christma',
'christmas',
'christmasparti',
'christmastre',
'christoph',
'christopherlambert',
'christopherlloyd',
'christophernolan',
'christopherplumm',
'christopherwalken',
'christophwaltz',
'chrisweitz',
'chronicl',
'chuck',
'church',
'cia',
'ciaránhind',
'cigarettesmok',
'cillianmurphi',
'cinema',
'circl',
'circu',
'circuit',
'circumst',
'citi',
'citizen',
'city',
'civil',
'civilian',
'civilwar',
'claim',
'clair',
'clairedan',
'claireforlani',
'clan',
'clark',

'clash',
'class',
'classdiffer',
'classic',
'classmat',
'classroom',
'claudevandamm',
'clay',
'clean',
'clear',
'clerk',
'clever',
'client',
'clients',
'cliff',
'climat',
'climb',
'clinteastwood',
'cliveowen',
'clock',
'clone',
'close',
'closer',
'cloud',
'clown',
'club',
'clue',
'clueless',
'clutch',
'coach',
'coast',
'cocain',
'code',
'coffin',
'cohen',
'col',
'cold',
'coldwar',
'cole',
'colin',
'colinfarrel',
'colinfirth',
'collaps',
'colleagu',
'collect',

'collector',
'colleg',
'college',
'collid',
'collis',
'colombia',
'colonel',
'coloni',
'color',
'colorado',
'coma',
'combat',
'combin',
'come',
'comeback',
'comed',
'comedi',
'comedian',
'comedy',
'comet',
'comfort',
'comic',
'coming',
'comingofag',
'comingout',
'command',
'commando',
'commerci',
'commiss',
'commit',
'common',
'commun',
'communist',
'community',
'compani',
'companion',
'company',
'compet',
'competit',
'competition',
'complet',
'complex',
'complic',
'compos',
'compuls',

'comput',
'computerviru',
'conan',
'concern',
'concert',
'concoct',
'condit',
'condition',
'conduct',
'confeder',
'confess',
'confid',
'confin',
'conflict',
'confront',
'confus',
'congress',
'conman',
'connect',
'connecticut',
'connel',
'connor',
'conquer',
'consequ',
'consequences',
'conserv',
'consid',
'conspir',
'conspiraci',
'conspiracy',
'constant',
'constantli',
'construct',
'consum',
'contact',
'contain',
'contemporari',
'contend',
'content',
'contest',
'contin',
'contract',
'contractor',
'control',
'controversi',

```
'convent',
'converg',
'convers',
'convict',
'convinc',
'cook',
'cool',
'cooper',
'cop',
'cope',
...]
```

```
In [74]: ['loved', 'love', 'loving']
```

```
Out[74]: ['loved', 'love', 'loving']
```

```
In [76]: ps.stem('loving')
```

```
Out[76]: 'love'
```

```
In [85]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [87]: #check the similiarity score of first with all
#other movies like haar movie ka haar ke sath distance find out krna hai similiarity nikalni hai , 1 means similiarity z
#0 means similiarity kaam hai
similarity = cosine_similarity(vectors)
```

```
In [88]: similarity
```

```
Out[88]: array([[1.          , 0.08346223, 0.0860309 , ..., 0.04499213, 0.          ,
        0.          ],
       [0.08346223, 1.          , 0.06063391, ..., 0.02378257, 0.          ,
        0.02615329],
       [0.0860309 , 0.06063391, 1.          , ..., 0.02451452, 0.          ,
        0.          ],
       ...,
       [0.04499213, 0.02378257, 0.02451452, ..., 1.          , 0.03962144,
        0.04229549],
       [0.          , 0.          , 0.          , ..., 0.03962144, 1.          ,
        0.08714204],
       [0.          , 0.02615329, 0.          , ..., 0.04229549, 0.08714204,
        1.          ]])
```

```
In [89]: new_df[new_df['title'] == 'Batman Begins'].index[0]
```

```
Out[89]: 119
```

```
In [ ]: # first jab movie find out krne ke liye firts have to find out index,
# index = new_df[new_df['title'] == 'batman begins'].index[0]
#distance nikalne hain
#then sort distances=> distances was in array
#sorted(similarity[0])
#distances = sorted(similarity[index])
```

```
In [90]: movie_index = new_df[new_df['title'] == 'Batman Begins'].index[0]
```

```
In [91]: movie_index
```

```
Out[91]: 119
```

```
In [94]: sorted(similarity[0][-10:-1])
```

```
Out[94]: [0.0,
0.0,
0.0,
0.0,
0.019252140716412975,
0.04223886030955117,
0.04499212706658476,
0.046829290579084706,
0.052631578947368425]
```

```
In [98]: #list(enumerate(similarity[0]))
#sorted(list(enumerate(similarity[0])),reverse=True ,key = lambda x: x[1])
sorted(list(enumerate(similarity[0])),reverse=True ,key = lambda x: x[1])[1:6]
```

```
Out[98]: [(1216, 0.28676966733820225),
(2409, 0.26901379342448517),
(3730, 0.2605130246476754),
(507, 0.255608593705383),
(539, 0.2503866978335957)]
```

```
In [101... 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 x: x[1])[1:6]
```

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

In [105...

```
recommend('Avatar')
```

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

In [103...

```
new_df.iloc[1216].title
```

Out[103]:

```
'Aliens vs Predator: Requiem'
```

In []: