

In [3]: `import numpy as np`

```
In [4]: import pandas as pd
```

```
In [5]: movies=pd.read_csv('tmdb_5000_movies.csv')
```

```
In [6]: credits=pd.read_csv('tmdb_5000_credits.csv')
```

```
In [7]: movies.head(1)
```

Out[7]:

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

Out[8]: (4803, 20)

```
In [9]: credits.head(1)
```

Out[9]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "order": 1}, {"cast_id": 243, "character": "Neelix", "order": 2}, {"cast_id": 244, "character": "Miles Quaritch", "order": 3}, {"cast_id": 245, "character": "Moat", "order": 4}, {"cast_id": 246, "character": "Pond", "order": 5}, {"cast_id": 247, "character": "Tug", "order": 6}, {"cast_id": 248, "character": "Woolie", "order": 7}, {"cast_id": 249, "character": "Dug", "order": 8}, {"cast_id": 250, "character": "Gill", "order": 9}, {"cast_id": 251, "character": "Kip", "order": 10}, {"cast_id": 252, "character": "Ley", "order": 11}, {"cast_id": 253, "character": "Munk", "order": 12}, {"cast_id": 254, "character": "Neytiri", "order": 13}, {"cast_id": 255, "character": "Oo", "order": 14}, {"cast_id": 256, "character": "Pez", "order": 15}, {"cast_id": 257, "character": "Rani", "order": 16}, {"cast_id": 258, "character": "Sami", "order": 17}, {"cast_id": 259, "character": "Tay", "order": 18}, {"cast_id": 260, "character": "Viper", "order": 19}, {"cast_id": 261, "character": "Wey", "order": 20}, {"cast_id": 262, "character": "Zit", "order": 21}], [{"id": 1, "name": "Cameron", "role": "Director"}, {"id": 2, "name": "Stephanie", "role": "Production Design"}, {"id": 3, "name": "James", "role": "Visual Effects"}, {"id": 4, "name": "Ed", "role": "Cinematography"}, {"id": 5, "name": "John", "role": "Sound Mixing"}, {"id": 6, "name": "Mark", "role": "Visual Effects"}, {"id": 7, "name": "Hans", "role": "Music Score"}, {"id": 8, "name": "Caterina", "role": "Production Design"}, {"id": 9, "name": "Caterina", "role": "Production Design"}]	

```
In [10]: credits.shape
```

Out[10]: (4803, 4)

```
In [11]: print(type(credits))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
In [12]: movies = movies.merge(credits, on='title')
```

```
In [13]: movies.shape
```

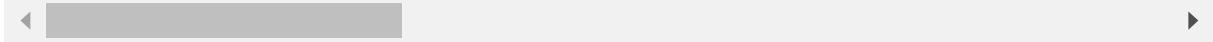
Out[13]: (4809, 23)

In [14]: movies.head()

Out[14]:

	budget	genres	homepage	id	keywords	original_
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...}	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...]	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...}	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...]	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...}	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...]	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...}	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, ...]	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...}	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": ...]	

5 rows × 23 columns



In [15]: 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 [16]: #genre

```
#id
#keywords
#title
#overview
#cast
#crew
```

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

In [17]: `movies.head()`

Out[17]:

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



In [18]: `movies.shape`

Out[18]: (4809, 7)

In [19]: `movies.isnull().sum()`

Out[19]:

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

In [20]: `movies.dropna(inplace=True)`

```
In [21]: movies.duplicated().sum()
```

```
Out[21]: 0
```

```
In [22]: movies.iloc[0].genres
```

```
Out[22]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [23]: #'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'  
#['Actions', 'Adventure', 'Fantasy', 'Sciencefi']
```

```
In [24]: import ast #for converting to list
```

```
In [25]: def convert(obj):  
    L = []  
  
    for i in ast.literal_eval(obj):  
        L.append(i['name'])  
  
    return L
```

```
In [26]: movies.dropna(inplace=True)
```

```
In [27]: movies['genres'] = movies['genres'].apply(convert)
```

In [28]: movies.head()

Out[28]:

	movie_id	title	overview	genres	keywords	cast	cre...
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id": ...}	[{"cast_id": 242, "character": "Jake Sully", "...}	{"credit_id": "52fe48009251416c750aca2", "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": "52fe4232c3a36847f800b57", "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": "54805967c3a36829b5002c4", "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": "52fe4781c3a36847f81398c", "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": "52fe479ac3a36847f813ea", "de...

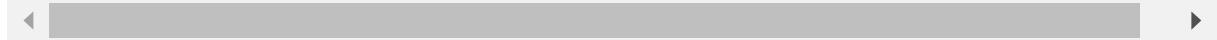


In [29]: movies['keywords']=movies['keywords'].apply(convert)

In [30]: movies.head()

Out[30]:

	movie_id	title	overview	genres	keywords	cast	cre...
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	{"cast_id": 242, "character": "Jake Sully", ...}	"52fe48009251416c750aca2... "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...	"52fe4232c3a36847f800b57... "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...	"54805967c3a36829b5002c4... "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..."}	"52fe4781c3a36847f81398c... "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..."}	"52fe479ac3a36847f813eaai... "de



```
In [31]: movies['cast'][0]
```

```
Out[31]: '[{"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\Tey", "credit_id": "52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso", "order": 9}, {"cast_id": 12, "character": "Dr. Max Patel", "credit_id": "52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "name": "Dileep Rao", "order": 10}, {"cast_id": 13, "character": "Lyle Wainfleet", "credit_id": "52fe48009251416c750ac9f5", "gender": 2, "id": 98215, "name": "Matt Gerald", "order": 11}, {"cast_id": 32, "character": "Private Fike", "credit_id": "52fe48009251416c750aca5b", "gender": 2, "id": 154153, "name": "Sean Anthony Mora n", "order": 12}, {"cast_id": 33, "character": "Cryo Vault Med Tech", "credit_id": "52fe48009251416c750aca5f", "gender": 2, "id": 397312, "name": "Jason WhYTE", "order": 13}, {"cast_id": 34, "character": "Venture Star Crew Chief", "credit_id": "52fe48009251416c750aca63", "gender": 2, "id": 42317, "name": "Scott Lawrence", "order": 14}, {"cast_id": 35, "character": "Lock Up Trooper", "credit_id": "52fe48009251416c750aca67", "gender": 2, "id": 986734, "name": "Kelly Kilgour", "order": 15}, {"cast_id": 36, "character": "Shuttle Pilot", "credit_id": "52fe48009251416c750aca6b", "gender": 0, "id": 1207227, "name": "James Patrick Pitt", "order": 16}, {"cast_id": 37, "character": "Shuttle Co-Pilot", "credit_id": "52fe48009251416c750aca6f", "gender": 0, "id": 1180936, "name": "Sean Patrick Murphy", "order": 17}, {"cast_id": 38, "character": "Shuttle Crew Chief", "credit_id": "52fe48009251416c750aca73", "gender": 2, "id": 1019578, "name": "Peter Dillon", "order": 18}, {"cast_id": 39, "character": "Tractor Operator / Troupe", "credit_id": "52fe48009251416c750aca77", "gender": 0, "id": 91443, "name": "Kevin Dorman", "order": 19}, {"cast_id": 40, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7b", "gender": 2, "id": 173391, "name": "Kelson Henderson", "order": 20}, {"cast_id": 41, "character": "Dragon Gunship Gunner", "credit_id": "52fe48009251416c750aca7f", "gender": 0, "id": 1207236, "name": "David Van Horn", "order": 21}, {"cast_id": 42, "character": "Dragon Gunship Navigator", "credit_id": "52fe48009251416c750aca83", "gender": 0, "id": 215913, "name": "Jacob Tomuri", "order": 22}, {"cast_id": 43, "character": "Suit #1", "credit_id": "52fe48009251416c750aca87", "gender": 0, "id": 143206, "name": "Michael Blain-Rozgay", "order": 23}, {"cast_id": 44, "character": "Suit #2", "credit_id": "52fe48009251416c750aca8b", "gender": 2, "id": 169676, "name": "Jon Curry", "order": 24}, {"cast_id": 46, "character": "Ambient Room Tech", "credit_id": "52fe48009251416c750aca8f", "gender": 0, "id": 1048610, "name": "Luke Hawker", "order": 25}, {"cast_id": 47, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48009251416c750aca93", "gender": 0, "id": 42288, "name": "Woody Schultz", "order": 26}, {"cast_id": 48, "character": "Horse Clan Leader", "credit_id": "52fe48009251416c750aca97", "gender": 2, "id": 68278, "name": "Peter Mensah", "order": 27}, {"cast_id": 49, "character": "Link Room Tech", "credit_id": "52fe48009251416c750aca9d", "gender": 0, "id": 1048611, "name": "Liam Neeson"}]
```

d": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247, "name": "Sonia Ye e", "order": 28}, {"cast_id": 50, "character": "Basketball Avatar / Troupe", "credit_id": "52fe48009251416c750aca9f", "gender": 1, "id": 1207248, "name": "Jahnel Curfman", "order": 29}, {"cast_id": 51, "character": "Basketball Avat ar", "credit_id": "52fe48009251416c750acaa3", "gender": 0, "id": 89714, "nam e": "Ilram Choi", "order": 30}, {"cast_id": 52, "character": "Na\'vi Child", "credit_id": "52fe48009251416c750acaa7", "gender": 0, "id": 1207249, "name": "Kyla Warren", "order": 31}, {"cast_id": 53, "character": "Troupe", "credit_i d": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Rou main", "order": 32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52f e48009251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "or der": 33}, {"cast_id": 57, "character": "Troupe", "credit_id": "52fe480092514 16c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 55, "character": "Troupe", "credit_id": "52fe48009251416c750acab 3", "gender": 0, "id": 1207251, "name": "Taylor Kibby", "order": 35}, {"cast_ id": 56, "character": "Troupe", "credit_id": "52fe48009251416c750acab7", "gen der": 0, "id": 1207252, "name": "Jodie Landau", "order": 36}, {"cast_id": 58, "character": "Troupe", "credit_id": "52fe48009251416c750acabf", "gender": 0, "id": 1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "characte r": "Troupe", "credit_id": "52fe48009251416c750acac3", "gender": 0, "id": 120 7257, "name": "Cullen B. Madden", "order": 38}, {"cast_id": 60, "character": "Troupe", "credit_id": "52fe48009251416c750acac7", "gender": 0, "id": 120725 9, "name": "Joseph Brady Madden", "order": 39}, {"cast_id": 61, "character": "Troupe", "credit_id": "52fe48009251416c750acacb", "gender": 0, "id": 120726 2, "name": "Frankie Torres", "order": 40}, {"cast_id": 62, "character": "Troup e", "credit_id": "52fe48009251416c750acacf", "gender": 1, "id": 1158600, "na me": "Austin Wilson", "order": 41}, {"cast_id": 63, "character": "Troupe", "c redit_id": "52fe48019251416c750acad3", "gender": 1, "id": 983705, "name": "Sa ra Wilson", "order": 42}, {"cast_id": 64, "character": "Troupe", "credit_id": "52fe48019251416c750acad7", "gender": 0, "id": 1207263, "name": "Tamica Washi ngton-Miller", "order": 43}, {"cast_id": 65, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acadb", "gender": 1, "id": 1145098, "name": "Lucy Briant", "order": 44}, {"cast_id": 66, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acadf", "gender": 2, "id": 33305, "name": "N athan Meister", "order": 45}, {"cast_id": 67, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae3", "gender": 0, "id": 1207264, "name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae7", "gender": 2, "id": 33311, "name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "character": "Op Center St aff", "credit_id": "52fe48019251416c750acaeb", "gender": 0, "id": 1207265, "n ame": "Paul Yates", "order": 48}, {"cast_id": 70, "character": "Op Center Dut y Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 12072 66, "name": "Wray Wilson", "order": 49}, {"cast_id": 71, "character": "Op Cen ter Staff", "credit_id": "52fe48019251416c750acaf3", "gender": 2, "id": 5449 2, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "character": "Danc er", "credit_id": "52fe48019251416c750acaf7", "gender": 0, "id": 1207267, "nam e": "Melvin Leno Clark III", "order": 51}, {"cast_id": 73, "character": "Danc er", "credit_id": "52fe48019251416c750acafb", "gender": 0, "id": 1207268, "na me": "Carvon Futrell", "order": 52}, {"cast_id": 74, "character": "Dancer", "credit_id": "52fe48019251416c750acaff", "gender": 0, "id": 1207269, "name": "Brandon Jelkes", "order": 53}, {"cast_id": 75, "character": "Dancer", "credi t_id": "52fe48019251416c750acb03", "gender": 0, "id": 1207270, "name": "Micah Moch", "order": 54}, {"cast_id": 76, "character": "Dancer", "credit_id": "52f e48019251416c750acb07", "gender": 0, "id": 1207271, "name": "Hanniyah Muhamma d", "order": 55}, {"cast_id": 77, "character": "Dancer", "credit_id": "52fe48 019251416c750acb0b", "gender": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "credit_id": "52fe480192

```
51416c750acb0f", "gender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_id": 79, "character": "Dancer", "credit_id": "52fe48019251416c750acb13", "gender": 0, "id": 1207274, "name": "April Marie Thomas", "order": 58}, {"cast_id": 80, "character": "Dancer", "credit_id": "52fe48019251416c750acb17", "gender": 0, "id": 1207275, "name": "Bravita A. Threatt", "order": 59}, {"cast_id": 81, "character": "Mining Chief (uncredited)", "credit_id": "52fe48019251416c750acb1b", "gender": 0, "id": 1207276, "name": "Colin Blasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (uncredited)", "credit_id": "52fe48019251416c750acb1f", "gender": 0, "id": 107969, "name": "Mike Bodnar", "order": 61}, {"cast_id": 83, "character": "Richard (uncrated)", "credit_id": "52fe48019251416c750acb23", "gender": 0, "id": 1207278, "name": "Matt Clayton", "order": 62}, {"cast_id": 84, "character": "Nav'i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gender": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "order": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2, "id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 89, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 90, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 91, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 92, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 93, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 94, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 95, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 96, "character": "Ops Centreworker (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 97, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 98, "character": "Female Marine (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name": "T. J. Storm", "order": 77}, {"cast_id": 99, "character": "Col. Quaritch's Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "Jodie Taylor", "order": 78}, {"cast_id": 100, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 101, "character": "Geologist (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207292, "name": "Richard Whiteside", "order": 80}, {"cast_id": 102, "character": "Na'vi (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Nikie Zambo", "order": 81}, {"cast_id": 103, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}]]
```

```
In [32]: def convert(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 [33]: movies['cast']=movies['cast'].apply(convert)
```

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

Out[34]:

	movie_id	title	overview	genres	keywords	cast	ci
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": "52fe48009251416c750aca", "c...
1	285	Pirates of the Caribbean: At World's End	Captain Barbosa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...]	[Johnny Depp, Orlando Bloom, Keira Knightley]	{"credit_id": "52fe4232c3a36847f800b5", "c...
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": "54805967c3a36829b5002c", "c...
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": "52fe4781c3a36847f81398", "c...
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": "52fe479ac3a36847f813ea", "c...



```
In [35]: movies['crew'][0]
```

```
Out[35]: '[{"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": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gender": 0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "539c4a4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "5544ee3b925141499f0008fc", "department": "Sound", "gender": 2, "id": 1729, "job": "Original Music Composer", "name": "James Horner"}, {"credit_id": "52fe48009251416c750ac9c3", "department": "Directing", "gender": 2, "id": 2710, "job": "Director", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "gender": 2, "id": 2710, "job": "Writer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "department": "Editing", "gender": 2, "id": 2710, "job": "Editor", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca29", "department": "Production", "gender": 2, "id": 2710, "job": "Producer", "name": "James Cameron"}]
```

```
In [36]: def fetch_director(obj):
    L=[]
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            L.append(i['name'])
    return L
```

```
In [37]: movies['crew']=movies['crew'].apply(fetch_director)
```

In [38]: `movies.head()`

Out[38]:

	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 Barbosa, 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 [39]: `movies['overview'][0]`

Out[39]: '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 [40]: `movies['overview']=movies['overview'].apply(lambda x:x.split())`

In [41]: movies.head()

Out[41]:

	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 [42]: #'Sam Worthington'
#'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 [43]: `movies.head()`

Out[43]:

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

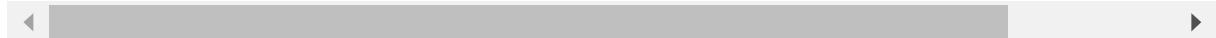
◀ ▶

In [44]: `movies['tags']=movies['overview']+movies['genres']+movies['keywords']+movies['`

In [45]: `movies.head()`

Out[45]:

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



In [46]: `new_df=movies[['movie_id', 'title', 'tags']]`

In [47]: `new_df['tags']=new_df['tags'].apply(lambda x: " ".join(x))`

```
C:\Users\DELL\AppData\Local\Temp\ipykernel_17032\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

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

In [48]: `new_df.head()`

Out[48]:

	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 Barbosa, 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 [49]: `import nltk`

In [50]: `from nltk.stem.porter import PorterStemmer
ps=PorterStemmer()`

In [51]: `def stem(text):
 y=[]`

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

        return " ".join(y)
```

In [52]: `new_df['tags']=new_df['tags'].apply(stem)`

C:\Users\DELL\AppData\Local\Temp\ipykernel_17032\3514595201.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 [53]: `new_df['tags'][0]`

Out[53]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniu mission, but becom torn between follow order and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar space coloni societi spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'

In [54]: `new_df['tags']=new_df['tags'].apply(lambda x:x.lower())`

```
C:\Users\DELL\AppData\Local\Temp\ipykernel_17032\4224080999.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 [55]: `new_df.head()`

Out[55]:

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

In [56]: `from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000,stop_words='english')`

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

In [58]: `vectors[0]`

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

In [59]: `cv.get_feature_names_out()`

Out[59]: `array(['000', '007', '10', ..., 'zone', 'zoo', 'zooeydeschanel'],
 dtype=object)`

In [60]: `stem('In the 22nd century, a paraplegic Marine is dispatched to the moon Pandor...`

Out[60]: `'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniu mission, but becom torn between follow order and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar space coloni societi spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'`

In [61]: `from sklearn.metrics.pairwise import cosine_similarity`

```
In [62]: similarity=cosine_similarity(vectors)
```

```
In [63]: def recommend(movie):
    movies_index=new_df[new_df['title']==movie].index[0]
    distances= similarity[movies_index]
    movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])
    for i in movies_list:
        print(new_df.iloc[i[0]].title)
```

```
In [64]: recommend('Avatar') #gives you 5 movie on the basis of genre
```

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

```
In [65]: import pickle
```

```
In [66]: pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

```
In [67]: pickle.dump(similarity,open('similarity.pkl','wb'))
```