

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

```
In [281]: 1 credit=pd.read_csv('tmdb_5000_credits.csv')
```

```
In [282]: 1 movies=pd.read_csv('tmdb_5000_movies.csv')
```

```
In [283]: 1 credit.head()
```

Out[283]:

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

```
In [284]: 1 movies.head()
```

Out[284]:

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

```
In [285]: 1 movies=movies.merge(credit,on='title') # Merging the two data set
```

```
In [286]: 1 movies.shape
```

Out[286]: (4809, 23)

```
In [287]: 1 movies.head(1)
```

Out[287]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...]	en	Avatar	In the 22nd century, a paraplegic Marine is di...

1 rows × 23 columns

```
In [288]: 1 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 [289]:

```
1 # Most importatant column
2 #genres
3 #id
4 #keywords
5 #title
6 #overview
7 #cast
8 #crew
9
10 movies=movies[['id','genres','keywords','title','overview','cast','crew']]
```

In [290]: 1 movies

Out[290]:

		id	genres	keywords	title	overview	cast	crew
0	19995		[{"id": 28, "name": "Action"}, {"id": 12, "name...}]	[{"id": 1463, "name": "culture clash"}, {"id": ...}]	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"cast_id": 242, "character": "Jake Sully", "...}]	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285		[{"id": 12, "name": "Adventure"}, {"id": 14, "...}]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...}]	Pirates of the Caribbean: At World's End	Captain Barbosa, long believed to be dead, ha...	[{"cast_id": 4, "character": "Captain Jack Spa...}]	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647		[{"id": 28, "name": "Action"}, {"id": 12, "name...}]	[{"id": 470, "name": "spy"}, {"id": 818, "name...}]	Spectre	A cryptic message from Bond's past sends him o...	[{"cast_id": 1, "character": "James Bond", "cr...}]	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026		[{"id": 28, "name": "Action"}, {"id": 80, "nam...}]	[{"id": 849, "name": "dc comics"}, {"id": 853,...}]	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...}]	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529		[{"id": 28, "name": "Action"}, {"id": 12, "nam...}]	[{"id": 818, "name": "based on novel"}, {"id": ...}]	John Carter	John Carter is a war-weary, former military ca...	[{"cast_id": 5, "character": "John Carter", "c...}]	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...
...
4804	9367		[{"id": 28, "name": "Action"}, {"id": 80, "nam...}]	[{"id": 5616, "name": "united states\u2013mexi...}]	El Mariachi	El Mariachi just wants to play his guitar and ...	[{"cast_id": 1, "character": "El Mariachi", "...}]	[{"credit_id": "52fe44eec3a36847f80b280b", "de...
4805	72766		[{"id": 35, "name": "Comedy"}, {"id": 10749, "...}]	[]	Newlyweds	A newlywed couple's honeymoon is upended by th...	[{"cast_id": 1, "character": "Buzzy", "credit_...}]	[{"credit_id": "52fe487dc3a368484e0fb013", "de...
4806	231617		[{"id": 35, "name": "Comedy"}, {"id": 18, "nam...}]	[{"id": 248, "name": "date"}, {"id": 699, "nam...}]	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...	[{"cast_id": 8, "character": "Oliver O\u2019To...}]	[{"credit_id": "52fe4df3c3a36847f8275ecf", "de...
4807	126186		[]	[]	Shanghai Calling	When ambitious New York attorney Sam is sent t...	[{"cast_id": 3, "character": "Sam", "credit_id...}]	[{"credit_id": "52fe4ad9c3a368484e16a36b", "de...
4808	25975		[{"id": 99, "name": "Documentary"}]	[{"id": 1523, "name": "obsession"}, {"id": 224...}]	My Date with Drew	Ever since the second grade when he first saw ...	[{"cast_id": 3, "character": "Herself", "credi...}]	[{"credit_id": "58ce021b9251415a390165d9", "de...

4809 rows × 7 columns

```
In [291]: 1 movies.isnull().sum() # overview contain null value s
```

```
Out[291]: id      0  
genres    0  
keywords  0  
title     0  
overview  3  
cast      0  
crew      0  
dtype: int64
```

```
In [292]: 1 movies.dropna(inplace=True)
```

```
In [293]: 1 movies.isnull().sum()
```

```
Out[293]: id      0  
genres    0  
keywords  0  
title     0  
overview  0  
cast      0  
crew      0  
dtype: int64
```

```
In [294]: 1 movies['genres'][0]
```

```
Out[294]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [295]: 1 #to convert in ['Action', 'Adventure', 'Fantasy', 'Science fi']
```

```
In [296]: 1 import ast
```

```
In [297]: 1 def convert (obj):  
2     l=[]  
3     for i in ast.literal_eval(obj): # ast.literal eval convert the statement in string  
4         l.append(i['name'])  
5     return l
```

```
In [298]: 1 movies['genres']=movies['genres'].apply(convert)
```

```
In [299]: 1 #also onvert keywords
```

```
In [300]: 1 movies['keywords']=movies['keywords'].apply(convert)
```

In [301]: 1 movies.head()

Out[301]:

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

```
In [302]: 1 movies['cast'][0]
```

Out[302]: '[{"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": "CH 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": "52fe48009251416c750aca9f5", "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 Moran", "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": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247, "name": "Sonia Yee", "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 Avatar", "credit_id": "52fe48009251416c750acaaa3", "gender": 0, "id": 89714, "name": "Ilram Choi", "order": 30}, {"cast_id": 52, "character": "Na'vi Child", "credit_id": "52fe48009251416c750aca7", "gender": 0, "id": 1207249, "name": "Kyla Warren", "order": 31}, {"cast_id": 53, "character": "Troupe", "credit_id": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Roumain", "order": 32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52fe48009251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id": 55, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 56, "character": "Troupe", "credit_id": "52fe48009251416c750acab3", "gender": 0, "id": 1207251, "name": "Taylor Kibby", "order": 35}, {"cast_id": 56, "character": "Troupe", "credit_id": "52fe48009251416c750acab7", "gender": 0, "id": 1207252, "name": "Jodie Landau", "order": 36}'

r": 36}, {"cast_id": 58, "character": "Troupe", "credit_id": "52fe48009251416c750acabf", "gender": 0, "id": 1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "character": "Troupe", "credit_id": "52fe48009251416c750acac3", "gender": 0, "id": 1207257, "name": "Cullen B. Madden", "order": 38}, {"cast_id": 60, "character": "Troupe", "credit_id": "52fe48009251416c750acac7", "gender": 0, "id": 1207259, "name": "Joseph Brady Madden", "order": 39}, {"cast_id": 61, "character": "Troupe", "credit_id": "52fe48009251416c750acacb", "gender": 0, "id": 1207262, "name": "Frankie Torres", "order": 40}, {"cast_id": 62, "character": "Troupe", "credit_id": "52fe48009251416c750acacf", "gender": 1, "id": 1158600, "name": "Austin Wilson", "order": 41}, {"cast_id": 63, "character": "Troupe", "credit_id": "52fe48019251416c750acad3", "gender": 1, "id": 983705, "name": "Sara Wilson", "order": 42}, {"cast_id": 64, "character": "Troupe", "credit_id": "52fe48019251416c750acad7", "gender": 0, "id": 1207263, "name": "Tamica Washington-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": "Nathan Meister", "order": 45}, {"cast_id": 67, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 1207264, "name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaef7", "gender": 2, "id": 33311, "name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaeb", "gender": 0, "id": 1207265, "name": "Paul Yates", "order": 48}, {"cast_id": 70, "character": "Op Center Duty Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 1207266, "name": "Wray Wilson", "order": 49}, {"cast_id": 71, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaef3", "gender": 2, "id": 54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "character": "Dancer", "credit_id": "52fe48019251416c750acaf7", "gender": 0, "id": 1207267, "name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 73, "character": "Dancer", "credit_id": "52fe48019251416c750acaf3", "gender": 0, "id": 1207268, "name": "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", "credit_id": "52fe48019251416c750acb03", "gender": 0, "id": 1207270, "name": "Micah Moch", "order": 54}, {"cast_id": 76, "character": "Dancer", "credit_id": "52fe48019251416c750acb07", "gender": 0, "id": 1207271, "name": "Hanniyah Muhammad", "order": 55}, {"cast_id": 77, "character": "Dancer", "credit_id": "52fe48019251416c750acb0b", "gender": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "credit_id": "52fe48019251416c750acb0f", "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 Bleasdale", "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 (uncredited)", "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": 14, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 90, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 91, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 92, "character": "Last Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 93, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Vindum Raah (uncredited)"}

```
e": "Stuart Pollock", "order": 73}, {"cast_id": 94, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 95, "character": "Ops Centreworker (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 96, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 97, "character": "Col. Quaritch\`s Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "T. J. Storm", "order": 77}, {"cast_id": 98, "character": "Female Marine (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name": "Jodie Taylor", "order": 78}, {"cast_id": 99, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 100, "character": "Geologist (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207292, "name": "Richard Whiteside", "order": 80}, {"cast_id": 101, "character": "Na\'vi (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Nikie Zambo", "order": 81}, {"cast_id": 102, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}]]
```

```
In [303]: 1 def convert3 (obj):
2     l=[]
3     counter=0
4     for i in ast.literal_eval(obj):# ast.literal eval convert the statement in string
5         if counter!=3: # we set a condition need only 3 name actor name from cast
6             l.append(i['name'])
7             counter+=1
8         else:
9             break
10    return l
```

```
In [304]: 1 movies['cast']=movies['cast'].apply(convert3)
```

```
In [305]: 1 movies.head()
```

Out[305]:

	id	genres	keywords	title	overview	cast	crew
0	19995	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...]	Avatar	In the 22nd century, a paraplegic Marine is di...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[{"credit_id": "52fe48009251416c750aca23", "de..."]
1	285	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...]	Pirates of the Caribbean: At World's End	Captain Barbosa, long believed to be dead, ha...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[{"credit_id": "52fe4232c3a36847f800b579", "de..."]
2	206647	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...]	Spectre	A cryptic message from Bond's past sends him o...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[{"credit_id": "54805967c3a36829b5002c41", "de..."]
3	49026	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...]	The Dark Knight Rises	Following the death of District Attorney Harve...	[Christian Bale, Michael Caine, Gary Oldman]	[{"credit_id": "52fe4781c3a36847f81398c3", "de..."]
4	49529	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...]	John Carter	John Carter is a war-weary, former military ca...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[{"credit_id": "52fe479ac3a36847f813eaa3", "de..."]

```
In [306]: 1 movies['crew'][0]
```

```
Out[306]: [{"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"}, {"credit_id": "52fe48009251416c750aca3f", "department": "Writing", "gender": 2, "id": 2710, "job": "Screenplay", "name": "James Cameron"}, {"credit_id": "539c4987c3a36810ba0021a4", "department": "Art", "gender": 2, "id": 7236, "job": "Art Direction", "name": "James Cameron"}]
```

```
In [307]: 1 def convert4 (obj):
```

```
2     l=[]
3     for i in ast.literal_eval(obj): # ast.literal_eval convert the statement in string
4         if i['job']=='Director': # we need only director name
5             l.append(i['name'])
6     return l
```

```
In [308]: 1 movies['crew']=movies['crew'].apply(convert4)
```

```
In [309]: 1 movies.head()
```

```
Out[309]:
```

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

```
In [310]: 1 movies['overview'] # this column in string formate So we convert in to List formate because
          2 #and crew are in list formate
```

```
Out[310]: 0      In the 22nd century, a paraplegic Marine is di...
1      Captain Barbosa, long believed to be dead, ha...
2      A cryptic message from Bond's past sends him o...
3      Following the death of District Attorney Harve...
4      John Carter is a war-weary, former military ca...
...
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: overview, Length: 4806, dtype: object
```

```
In [311]: 1 movies['overview']=movies['overview'].apply(lambda x:x.split())
```

```
In [312]: 1 movies.head()
```

```
Out[312]:
```

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

```
In [313]: 1 def remove_space(string): # removing space from all col
          2 k=[]
          3 for i in string:
          4     j=i.replace(' ','')
          5     k.append(j)
          6 return k
```

```
In [314]: 1 movies['genres']=movies['genres'].apply(remove_space)
```

```
In [315]: 1 movies['cast']=movies['cast'].apply(remove_space)
```

```
In [316]: 1 movies['crew']=movies['crew'].apply(remove_space)
          2 movies['keywords']=movies['keywords'].apply(remove_space)
```

In [317]: 1 movies.head()

Out[317]:

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

In [318]: 1 movies['tags']=movies['overview']+movies['genres']+movies['keywords']+movies['cast']+movies['title']
2 # concat all column in tags

In [319]: 1 movies.head()

Out[319]:

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

```
In [320]: 1 new_df=movies[['id','title','tags']]
```

```
In [321]: 1 new_df.head()
```

Out[321]:

	id	title	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...

```
In [322]: 1 new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

C:\Users\USER\AppData\Local\Temp\ipykernel_2332\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 [324]: 1 new_df.head() # we check similarity based on tags
```

Out[324]:

	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 [325]: 1 # We convert your text tag in to vector
```

```
In [326]: 1 from sklearn.feature_extraction.text import CountVectorizer
```

```
In [327]: 1 cv=CountVectorizer(max_features=500,stop_words='english')
```

```
In [324]: 1 vector=cv.fit_transform(new_df['tags'])
```

```
In [336]: 1 vector.shape
```

Out[336]: (4806, 500)

```
In [338]: 1 # we compare the first movie vs 4806 movie then second movie to 4806 and so on till 4806
```

```
In [339]: 1 # we use cosine distance
```

```
In [341]: 1 from sklearn.metrics.pairwise import cosine_similarity
```

```
In [342]: 1 similarity =cosine_similarity(vector)
```

```
In [343]: 1 similarity.shape
```

```
Out[343]: (4806, 4806)
```

```
In [363]: 1 def recommend(movie):
2     # index fetch
3     movie_index=new_df[new_df['title']==movie].index[0]
4     distance=similarity[movie_index]
5     movie_list=sorted(list(enumerate(distance)),reverse=True,key=lambda x:x[1])[1:6]
6     for i in movie_list:
7         print(new_df.iloc[i[0]].title)
```

```
In [365]: 1 recommend('Batman Begins')
```

```
The Dark Knight
Batman & Robin
Amidst the Devil's Wings
Mi America
Dead Man Down
```

```
In [ ]: 1
```

In []: 1

In []: 1

```
In [ ]: 1
```

In []: 1

```
In [ ]: 1
```

In []: 1

In []: 1

In []: 1

```
In [ ]: 1
```

In []: 1

In []: 1

In []: 1

In []: 1

```
In [ ]: 1
```

In []: 1

In []: 1

```
In [ ]: 1
```

```
In [ ]: 1
```

```
In [ ]: 1
```

