

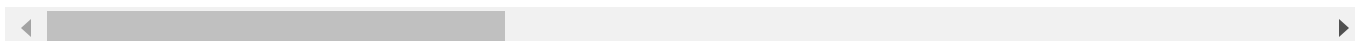

```
In [6]: # merging dataframes
```

```
In [7]: movies=movies.merge(credits,on='title')
movies[:2]
```

```
Out[7]:
```

	budget	genres	homepage	id	keywords	original_language	origi
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirati Ca Al

2 rows × 23 columns



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

```
Out[8]: (4809, 23)
```

```
In [9]: movies.columns
```

```
Out[9]: Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',  
              'original_title', 'overview', 'popularity', 'production_companies',  
              'production_countries', 'release_date', 'revenue', 'runtime',  
              'spoken_languages', 'status', 'tagline', 'title', 'vote_average',  
              'vote_count', 'movie_id', 'cast', 'crew'],  
              dtype='object')
```

```
In [10]: movies.isnull().sum()
```

```
keywords          0
original_language  0
original_title     0
overview          3
popularity         0
production_companies  0
production_countries  0
release_date       1
revenue            0
runtime            2
spoken_languages   0
status             0
tagline            844
title              0
vote_average       0
vote_count         0
movie_id           0
cast               0
crew               0
dtype: int64
```

```
In [11]: # choosing best fetures to recommend a movie
```

```
In [12]: df=movies[['movie_id','title','overview','genres','keywords','cast','crew']]
df.head()
```

```
Out[12]:
```

	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...

```
In [13]: #goal- movie_id+tags
```

```
In [14]: df['overview']
```

```
Out[14]: 0      In the 22nd century, a paraplegic Marine is di...
1      Captain Barbossa, long believed to be dead, ha...
2      A cryptic message from Bond's past sends him o...
3      Following the death of District Attorney Harve...
4      John Carter is a war-weary, former military ca...
...
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: overview, Length: 4809, dtype: object
```

```
In [15]: df['overview'][0]
```

```
Out[15]: '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 [16]: df['genres']
```

```
Out[16]: 0      [{"id": 28, "name": "Action"}, {"id": 12, "nam...
1      [{"id": 12, "name": "Adventure"}, {"id": 14, "...
2      [{"id": 28, "name": "Action"}, {"id": 12, "nam...
3      [{"id": 28, "name": "Action"}, {"id": 80, "nam...
4      [{"id": 28, "name": "Action"}, {"id": 12, "nam...
...
4804    [{"id": 28, "name": "Action"}, {"id": 80, "nam...
4805    [{"id": 35, "name": "Comedy"}, {"id": 10749, "...
4806    [{"id": 35, "name": "Comedy"}, {"id": 18, "nam...
4807                                           []
4808    [{"id": 99, "name": "Documentary"}]
Name: genres, Length: 4809, dtype: object
```

```
In [17]: df['genres'][0]
```

```
Out[17]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [18]: df['keywords']
```

```
Out[18]: 0      [{"id": 1463, "name": "culture clash"}, {"id": ...
1      [{"id": 270, "name": "ocean"}, {"id": 726, "na...
2      [{"id": 470, "name": "spy"}, {"id": 818, "name...
3      [{"id": 849, "name": "dc comics"}, {"id": 853,...
4      [{"id": 818, "name": "based on novel"}, {"id": ...

...
4804    [{"id": 5616, "name": "united states\u2013mexi...
4805                                           []
4806    [{"id": 248, "name": "date"}, {"id": 699, "nam...
4807                                           []
4808    [{"id": 1523, "name": "obsession"}, {"id": 224...
Name: keywords, Length: 4809, dtype: object
```

```
In [19]: df['keywords'][0]
```

```
Out[19]: '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386,
"name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "soci
ety"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"i
d": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "ali
en"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien planet"}, {"id": 10
987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldie
r"}, {"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165
431, "name": "anti war"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "n
ame": "mind and soul"}, {"id": 209714, "name": "3d"}]'
```

```
In [20]: df['cast'][0]
```

```
Out[20]: '[{"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 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": "52fe48009251416c750aca3", "gender": 0, "id": 89714, "name": "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_id": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Rou"}]
```

```
main": "order": 32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52fe4809251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id": 57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 55, "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}, {"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": "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 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": "52fe48019251416c750acaf3", "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": "52fe48019251416c750acafb", "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": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 0, "id": 1207282, "name": "Allan Henry", "order": 66}, {"cast_id": 88, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Allan Henry", "order": 67}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1207284, "name": "Allan Henry", "order": 68}, {"cast_id": 90, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 1207285, "name": "Allan Henry", "order": 69}, {"cast_id": 91, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207286, "name": "Allan Henry", "order": 70}, {"cast_id": 92, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207287, "name": "Allan Henry", "order": 71}, {"cast_id": 93, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207288, "name": "Allan Henry", "order": 72}, {"cast_id": 94, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 1207289, "name": "Allan Henry", "order": 73}, {"cast_id": 95, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Allan Henry", "order": 74}, {"cast_id": 96, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1207291, "name": "Allan Henry", "order": 75}, {"cast_id": 97, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 1207292, "name": "Allan Henry", "order": 76}, {"cast_id": 98, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207293, "name": "Allan Henry", "order": 77}, {"cast_id": 99, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 0, "id": 1207294, "name": "Allan Henry", "order": 78}, {"cast_id": 100, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207295, "name": "Allan Henry", "order": 79}, {"cast_id": 101, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 1207296, "name": "Allan Henry", "order": 80}, {"cast_id": 102, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb6f", "gender": 0, "id": 1207297, "name": "Allan Henry", "order": 81}, {"cast_id": 103, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb73", "gender": 0, "id": 1207298, "name": "Allan Henry", "order": 82}, {"cast_id": 104, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb77", "gender": 0, "id": 1207299, "name": "Allan Henry", "order": 83}, {"cast_id": 105, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb7b", "gender": 0, "id": 1207300, "name": "Allan Henry", "order": 84}, {"cast_id": 106, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb7f", "gender": 0, "id": 1207301, "name": "Allan Henry", "order": 85}, {"cast_id": 107, "character": "Trooper (uncredited)", "credit_id":
```



```
d": 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": 23696, "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": "Blast 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": "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 [21]: df['crew']
```

```
Out[21]: 0      [{"credit_id": "52fe48009251416c750aca23", "de...
1      [{"credit_id": "52fe4232c3a36847f800b579", "de...
2      [{"credit_id": "54805967c3a36829b5002c41", "de...
3      [{"credit_id": "52fe4781c3a36847f81398c3", "de...
4      [{"credit_id": "52fe479ac3a36847f813eaa3", "de...
...
4804     [{"credit_id": "52fe44eec3a36847f80b280b", "de...
4805     [{"credit_id": "52fe487dc3a368484e0fb013", "de...
4806     [{"credit_id": "52fe4df3c3a36847f8275ecf", "de...
4807     [{"credit_id": "52fe4ad9c3a368484e16a36b", "de...
4808     [{"credit_id": "58ce021b9251415a390165d9", "de...
Name: crew, Length: 4809, dtype: object
```

```
In [22]: df['crew'][0]
```

```
r": 0, "id": 1352962, "job": "Set Designer", "name": "Tammy S. Lee"}, {"credit_id": "5495a115c3a3680ff5002d71", "department": "Crew", "gender": 0, "id": 1357070, "job": "Transportation Coordinator", "name": "Denny Caira"}, {"credit_id": "5495a12f92514130fc002e94", "department": "Crew", "gender": 0, "id": 1357071, "job": "Transportation Coordinator", "name": "James Waitkus"}, {"credit_id": "5495976fc3a36811530026b0", "department": "Sound", "gender": 0, "id": 1360103, "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"}]
```

clear genres feature

```
In [23]: df['genres']
```

```
Out[23]: 0      [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
1      [{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
2      [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
3      [{"id": 28, "name": "Action"}, {"id": 80, "name": "Drama"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
4      [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
...
4804   [{"id": 28, "name": "Action"}, {"id": 80, "name": "Drama"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
4805   [{"id": 35, "name": "Comedy"}, {"id": 10749, "name": "TV-14"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
4806   [{"id": 35, "name": "Comedy"}, {"id": 18, "name": "Drama"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
4807   [{"id": 99, "name": "Documentary"}]
4808   [{"id": 99, "name": "Documentary"}]
Name: genres, Length: 4809, dtype: object
```

```
In [24]: df['genres'][0]
```

```
Out[24]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [25]: type(df['genres'][0])
```

```
Out[25]: str
```

```
In [26]: ast.literal_eval(df['genres'][0])# taking out dict data type from the string
```

```
Out[26]: [{'id': 28, 'name': 'Action'},
{'id': 12, 'name': 'Adventure'},
{'id': 14, 'name': 'Fantasy'},
{'id': 878, 'name': 'Science Fiction'}]
```

```
In [27]: for i in ast.literal_eval(df['genres'][0]):
        print (i['name'])
```

Action
Adventure
Fantasy
Science Fiction

```
In [28]: def fetch_genres(data):
        l=[]
        for i in ast.literal_eval(data):
            l.append(i['name'])

        return l
```

```
In [29]: fetch_genres(df['genres'][0])
```

Out[29]: ['Action', 'Adventure', 'Fantasy', 'Science Fiction']

```
In [30]: df['genres'].apply(fetch_genres)
```

Out[30]: 0 [Action, Adventure, Fantasy, Science Fiction]
1 [Adventure, Fantasy, Action]
2 [Action, Adventure, Crime]
3 [Action, Crime, Drama, Thriller]
4 [Action, Adventure, Science Fiction]

...
4804 [Action, Crime, Thriller]
4805 [Comedy, Romance]
4806 [Comedy, Drama, Romance, TV Movie]
4807 []
4808 [Documentary]
Name: genres, Length: 4809, dtype: object

```
In [31]: df['genres']=df['genres'].apply(fetch_genres)
```

```
In [32]: df[:2]
```

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]	[{"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...

keywords

```
In [33]: df['keywords']
```

```
Out[33]: 0      [{"id": 1463, "name": "culture clash"}, {"id": ...
1      [{"id": 270, "name": "ocean"}, {"id": 726, "na...
2      [{"id": 470, "name": "spy"}, {"id": 818, "name...
3      [{"id": 849, "name": "dc comics"}, {"id": 853,...
4      [{"id": 818, "name": "based on novel"}, {"id": ...

...
4804     [{"id": 5616, "name": "united states\u2013mexi...
4805                                           []
4806     [{"id": 248, "name": "date"}, {"id": 699, "nam...
4807                                           []
4808     [{"id": 1523, "name": "obsession"}, {"id": 224...
Name: keywords, Length: 4809, dtype: object
```

```
In [34]: df['keywords'][0]
```

```
Out[34]: '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386,
"name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "soci
ety"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"i
d": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "ali
en"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien planet"}, {"id": 10
987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldie
r"}, {"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165
431, "name": "anti war"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "n
ame": "mind and soul"}, {"id": 209714, "name": "3d"}]'
```

```
In [35]: def fetch_keywords(data):
l=[]
for i in ast.literal_eval(data):
    l.append(i['name'])

return l
```

```
In [36]: df['keywords']=df['keywords'].apply(fetch_keywords)
```

```
In [37]: df[:3]
```

```
Out[37]:
```

	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...

cast

```
In [38]: df['cast']
```

```
Out[38]: 0      [{"cast_id": 242, "character": "Jake Sully", "...
1      [{"cast_id": 4, "character": "Captain Jack Spa...
2      [{"cast_id": 1, "character": "James Bond", "cr...
3      [{"cast_id": 2, "character": "Bruce Wayne / Ba...
4      [{"cast_id": 5, "character": "John Carter", "c...

...
4804    [{"cast_id": 1, "character": "El Mariachi", "c...
4805    [{"cast_id": 1, "character": "Buzzy", "credit_...
4806    [{"cast_id": 8, "character": "Oliver O\u2019To...
4807    [{"cast_id": 3, "character": "Sam", "credit_id...
4808    [{"cast_id": 3, "character": "Herself", "credi...
Name: cast, Length: 4809, dtype: object
```

```
In [39]: df['cast'][0]
```

```
Out[39]: '[{"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 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": "52fe48009251416c750aca3", "gender": 0, "id": 89714, "name": "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_id": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Rou"}]
```

```
main", "order": 32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52fe4809251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id": 57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 55, "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}, {"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": "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 Staff", "credit_id": "52fe48019251416c750acae8", "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": "52fe48019251416c750acaf3", "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": "52fe48019251416c750acafb", "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": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 0, "id": 1207282, "name": "Allan Henry", "order": 66}, {"cast_id": 88, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Allan Henry", "order": 67}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1207284, "name": "Allan Henry", "order": 68}, {"cast_id": 90, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 1207285, "name": "Allan Henry", "order": 69}, {"cast_id": 91, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207286, "name": "Allan Henry", "order": 70}, {"cast_id": 92, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207287, "name": "Allan Henry", "order": 71}, {"cast_id": 93, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207288, "name": "Allan Henry", "order": 72}, {"cast_id": 94, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 1207289, "name": "Allan Henry", "order": 73}, {"cast_id": 95, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Allan Henry", "order": 74}, {"cast_id": 96, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1207291, "name": "Allan Henry", "order": 75}, {"cast_id": 97, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 1207292, "name": "Allan Henry", "order": 76}, {"cast_id": 98, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207293, "name": "Allan Henry", "order": 77}, {"cast_id": 99, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 0, "id": 1207294, "name": "Allan Henry", "order": 78}, {"cast_id": 100, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207295, "name": "Allan Henry", "order": 79}, {"cast_id": 101, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 1207296, "name": "Allan Henry", "order": 80}, {"cast_id": 102, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb6f", "gender": 0, "id": 1207297, "name": "Allan Henry", "order": 81}, {"cast_id": 103, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb73", "gender": 0, "id": 1207298, "name": "Allan Henry", "order": 82}, {"cast_id": 104, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb77", "gender": 0, "id": 1207299, "name": "Allan Henry", "order": 83}, {"cast_id": 105, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb7b", "gender": 0, "id": 1207300, "name": "Allan Henry", "order": 84}, {"cast_id": 106, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb7f", "gender": 0, "id": 1207301, "name": "Allan Henry", "order": 85}, {"cast_id": 107, "character": "Trooper (uncredited)", "credit_id
```



```
d": 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": 23696, "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": "Blast 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": "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 [40]: df['crew']
```

```
Out[40]: 0      [{"credit_id": "52fe48009251416c750aca23", "de...
1      [{"credit_id": "52fe4232c3a36847f800b579", "de...
2      [{"credit_id": "54805967c3a36829b5002c41", "de...
3      [{"credit_id": "52fe4781c3a36847f81398c3", "de...
4      [{"credit_id": "52fe479ac3a36847f813eaa3", "de...
...
4804     [{"credit_id": "52fe44eec3a36847f80b280b", "de...
4805     [{"credit_id": "52fe487dc3a368484e0fb013", "de...
4806     [{"credit_id": "52fe4df3c3a36847f8275ecf", "de...
4807     [{"credit_id": "52fe4ad9c3a368484e16a36b", "de...
4808     [{"credit_id": "58ce021b9251415a390165d9", "de...
Name: crew, Length: 4809, dtype: object
```

In [41]: df['crew'][0]

```
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": "5592b1bfc3a368775d0010e7", "department": "Crew", "gender": 0, "id": 1483228, "job": "CG Supervisor", "name": "Sebastian Marino"}, {"credit_id": "5592b2049251415df8001078", "department": "Crew", "gender": 0, "id": 1483229, "job": "CG Supervisor", "name": "Matthias Menz"}, {"credit_id": "5592b27b92514152d800136a", "department": "Crew", "gender": 0, "id": 1483230, "job": "CG Supervisor", "name": "Sergei Nevshupov"}, {"credit_id": "5592b2c3c3a36869e800003c", "department": "Crew", "gender": 0, "id": 1483231, "job": "CG Supervisor", "name": "Philippe Rebours"}, {"credit_id": "5592b317c3a36877470012af", "department": "Crew", "gender": 0, "id": 1483232, "job": "CG Supervisor", "name": "Michael Takarangi"}, {"credit_id": "5592b345c3a36877470012bb", "department": "Crew", "gender": 0, "id": 1483233, "job": "CG Supervisor", "name": "David Weitzberg"}, {"credit_id": "5592b37cc3a368775100113b", "department": "Crew", "gender": 0, "id": 1483234, "job": "CG Supervisor", "name": "Ben White"}, {"credit_id": "573c8e2f9251413f5d000094", "department": "Crew", "gender": 1, "id": 1621932, "job": "Stunts", "name": "Min Windle"}]
```

```
In [42]: def fetch_cast(data):
         l=[]
         for i in ast.literal_eval(data)[:3]:
             l.append(i['name'])
         return l
```

In [43]: fetch_cast(df['cast'][0])

Out[43]: ['Sam Worthington', 'Zoe Saldana', 'Sigourney Weaver']

In [44]: fetch_cast(df['cast'][455])

Out[44]: ['Mel Gibson', 'Julia Roberts', 'Patrick Stewart']

In [45]: df['cast']=df['cast'].apply(fetch_cast)

```
In [46]: df[:3]
```

Out[46]:

	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...

crew

```
In [47]: df['crew'][0]
18265, job : Director of Photography , name : Mauro Fiore }, { credit_id : 5449
194d0e0a26748f001b39", "department": "Art", "gender": 0, "id": 42281, "job": "Set De
signer", "name": "Scott Herbertson"}, {"credit_id": "52fe48009251416c750aca05", "dep
artment": "Crew", "gender": 0, "id": 42288, "job": "Stunts", "name": "Woody Schult
z"}, {"credit_id": "5592aefb92514152de0010f5", "department": "Costume & Make-Up", "g
ender": 0, "id": 29067, "job": "Makeup Artist", "name": "Linda DeVetta"}, {"credit_i
d": "5592afa492514152de00112c", "department": "Costume & Make-Up", "gender": 0, "i
d": 29067, "job": "Hairstylist", "name": "Linda DeVetta"}, {"credit_id": "54959ed592
514130fc002e5d", "department": "Camera", "gender": 2, "id": 33302, "job": "Camera Op
erator", "name": "Richard Bluck"}, {"credit_id": "539c4891c3a36810ba002147", "depart
ment": "Art", "gender": 2, "id": 33303, "job": "Art Direction", "name": "Simon Brigh
t"}, {"credit_id": "54959c069251417a81001f3a", "department": "Visual Effects", "gend
er": 0, "id": 113145, "job": "Visual Effects Supervisor", "name": "Richard Martin"},
{"credit_id": "54959a0dc3a3680ff5002c8d", "department": "Crew", "gender": 2, "id": 5
8188, "job": "Visual Effects Editor", "name": "Steve R. Moore"}, {"credit_id": "52fe
48009251416c750aca1d", "department": "Editing", "gender": 2, "id": 58871, "job": "Ed
itor", "name": "John Refoua"}, {"credit_id": "54491a4dc3a3680fc30018ca", "departmen
t": "Art", "gender": 0, "id": 92359, "job": "Set Designer", "name": "Karl J. Marti
n"}, {"credit_id": "52fe48009251416c750aca35", "department": "Camera", "gender": 1,
"id": 72201, "job": "Director of Photography", "name": "Chiling Lin"}, {"credit_id":
```

```
In [48]: def fetch_director(data):
    l=[]
    for i in ast.literal_eval(data):
        if i['job']=='Director':
            l.append(i['name'])
    return l
```

```
In [49]: df['crew']=df['crew'].apply(fetch_director)
```

```
In [50]: df[:5]
```

```
Out[50]:
```

	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]

overview

```
In [51]: df['overview']
```

```
Out[51]: 0      In the 22nd century, a paraplegic Marine is di...
1      Captain Barbossa, long believed to be dead, ha...
2      A cryptic message from Bond's past sends him o...
3      Following the death of District Attorney Harve...
4      John Carter is a war-weary, former military ca...
...
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: overview, Length: 4809, dtype: object
```

```
In [52]: df['overview'][0]
```

```
Out[52]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization.'
```

```
In [53]: df.isnull().sum()
```

```
Out[53]: movie_id    0
         title       0
         overview    3
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [54]: df.dropna(inplace=True)
```

```
In [55]: df['overview']=df['overview'].apply(lambda x: x.split())
```

```
In [56]: df['overview']
```

```
Out[56]: 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
```

creating tags

```
In [57]: df['tags']=df['overview']+df['genres']+df['keywords']+df['cast']+df['crew']
```

```
In [58]: df[:3]
```

```
Out[58]:
```

	movie_id	title	overview	genres	keywords	cast	crew	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]	[In, the, 22nd, century,, a, paraplegic, Marin...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]	[Captain, Barbossa,, long, believed, to, be, d...
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]	[A, cryptic, message, from, Bond's, past, send...

```
In [59]: print(df['tags'][0])
```

```
['In', 'the', '22nd', 'century,', 'a', 'paraplegic', 'Marine', 'is', 'dispatched', 't  
o', 'the', 'moon', 'Pandora', 'on', 'a', 'unique', 'mission,', 'but', 'becomes', 'tor  
n', 'between', 'following', 'orders', 'and', 'protecting', 'an', 'alien', 'civilizatio  
n.', 'Action', 'Adventure', 'Fantasy', 'Science Fiction', 'culture clash', 'future',  
'space war', 'space colony', 'society', 'space travel', 'futuristic', 'romance', 'spac  
e', 'alien', 'tribe', 'alien planet', 'cgi', 'marine', 'soldier', 'battle', 'love affa  
ir', 'anti war', 'power relations', 'mind and soul', '3d', 'Sam Worthington', 'Zoe Sal  
dana', 'Sigourney Weaver', 'James Cameron']
```

final dataframe

```
In [60]: data=df[['movie_id','title','tags']]  
data
```

```
Out[60]:
```

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

4806 rows × 3 columns

```
In [61]: data['tags']=data['tags'].apply(lambda x: [i.replace(" ","")for i in x])
```

```
In [62]: print(data['tags'][0])
```

```
['In', 'the', '22nd', 'century,', 'a', 'paraplegic', 'Marine', 'is', 'dispatched', 't  
o', 'the', 'moon', 'Pandora', 'on', 'a', 'unique', 'mission,', 'but', 'becomes', 'tor  
n', 'between', 'following', 'orders', 'and', 'protecting', 'an', 'alien', 'civilizatio  
n.', 'Action', 'Adventure', 'Fantasy', 'ScienceFiction', 'cultureclash', 'future', 'sp  
acewar', 'spacecolony', 'society', 'spacetravel', 'futuristic', 'romance', 'space', 'a  
lien', 'tribe', 'alienplanet', 'cgi', 'marine', 'soldier', 'battle', 'loveaffair', 'an  
tiwar', 'powerrelations', 'mindandsoul', '3d', 'SamWorthington', 'ZoeSaldana', 'Sigour  
neyWeaver', 'JamesCameron']
```

In [63]: data

Out[63]:

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

4806 rows × 3 columns

In [64]: data['tags']=data['tags'].apply(lambda x:" ".join(x))

In [65]: data

Out[65]:

	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...
...
4804	9367	El Mariachi	El Mariachi just wants to play his guitar and ...
4805	72766	Newlyweds	A newlywed couple's honeymoon is upended by th...
4806	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...
4807	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...
4808	25975	My Date with Drew	Ever since the second grade when he first saw ...

4806 rows × 3 columns

NLP

```
In [67]: ps=PorterStemmer()
def preprocess_text(text):
    new_text=[]
    for i in text.split():
        lower=i.lower()
        new_text.append(ps.stem(lower))
    return " ".join(new_text)
```

```
In [68]: preprocess_text(data['tags'][0])
```

```
Out[68]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu miss
ion, but becom torn between follow order and protect an alien civilization. action adv
entur fantasi sciencefict cultureclash futur spacewar spacecoloni societi spacetravel
futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiw
ar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'
```

```
In [69]: data['tags'].apply(preprocess_text)
```

```
Out[69]: 0      in the 22nd century, a parapleg marin is dispa...
1      captain barbossa, long believ to be dead, ha c...
2      a cryptic messag from bond' past send him on a...
3      follow the death of district attorney harvey d...
4      john carter is a war-weary, former militari ca...

...

4804    el mariachi just want to play hi guitar and ca...
4805    a newlyw couple' honeymoon is upend by the arr...
4806    "signed, sealed, delivered" introduc a dedic q...
4807    when ambiti new york attorney sam is sent to s...
4808    ever sinc the second grade when he first saw h...
Name: tags, Length: 4806, dtype: object
```

```
In [70]: data['tags']=data['tags'].apply(preprocess_text)
```

```
In [71]: data
```

```
Out[71]:
```

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond' past send him on a...
3	49026	The Dark Knight Rises	follow the death of district attorney harvey d...
4	49529	John Carter	john carter is a war-weary, former militari ca...
...
4804	9367	El Mariachi	el mariachi just want to play hi guitar and ca...
4805	72766	Newlyweds	a newlyw couple' honeymoon is upend by the arr...
4806	231617	Signed, Sealed, Delivered	"signed, sealed, delivered" introduc a dedic q...
4807	126186	Shanghai Calling	when ambiti new york attorney sam is sent to s...
4808	25975	My Date with Drew	ever sinc the second grade when he first saw h...

4806 rows × 3 columns

bow (bag of words)

```
In [72]: cv = CountVectorizer(max_features=5000, stop_words='english')
```

```
In [74]: cv
```

```
Out[74]: CountVectorizer(max_features=5000, stop_words='english')
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [75]: data['tags']
```

```
Out[75]: 0      in the 22nd century, a parapleg marin is dispa...
1      captain barbossa, long believ to be dead, ha c...
2      a cryptic messag from bond' past send him on a...
3      follow the death of district attorney harvey d...
4      john carter is a war-weary, former militari ca...
...
4804    el mariachi just want to play hi guitar and ca...
4805    a newlyw couple' honeymoon is upend by the arr...
4806    "signed, sealed, delivered" introduc a dedic q...
4807    when ambiti new york attorney sam is sent to s...
4808    ever sinc the second grade when he first saw h...
Name: tags, Length: 4806, dtype: object
```

```
In [77]: vectors= cv.fit_transform(data['tags'])
vectors
```

```
Out[77]: <4806x5000 sparse matrix of type '<class 'numpy.int64'>'
with 145432 stored elements in Compressed Sparse Row format>
```

```
In [78]: vectors =cv.fit_transform(data['tags']).toarray()
```

```
In [79]: vectors
```

```
Out[79]: 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 [80]: len(vectors)
```

```
Out[80]: 4806
```

```
In [81]: vectors.shape
```

```
Out[81]: (4806, 5000)
```

```
In [82]: cv.get_feature_names_out()
```

```
Out[82]: array(['000', '007', '10', ..., 'zone', 'zoo', 'zoeydeschanel'],  
              dtype=object)
```

```
In [83]: for i in cv.get_feature_names_out():  
         print(i)
```

```
000  
007  
10  
100  
11  
12  
13  
14  
15  
16  
17  
17th  
18  
18th  
18thcenturi  
19  
1910  
1920  
1930  
1940
```

calculate cosine_similarity between vectors

```
In [84]: vectors
```

```
Out[84]: 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 [91]: sorted(similarity[0],reverse=True)
```

```
Out[91]: [1.0000000000000002,  
0.28676966733820225,  
0.26901379342448517,  
0.2605130246476754,  
0.255608593705383,  
0.25038669783359574,  
0.24511108480187255,  
0.24455799402225922,  
0.23179316248638276,  
0.23174488732966075,  
0.2278389747471728,  
0.2252817784447915,  
0.22269966704152225,  
0.21853668936906193,  
0.21239769762143662,  
0.2108663315950723,  
0.2105263157894737,  
0.20443988269091456,  
0.20437977982832192,  
0.20305070136100075]
```

```
In [92]: list(enumerate(similarity[0]))
```

```
Out[92]: [(0, 1.0000000000000002),  
(1, 0.08346223261119858),  
(2, 0.08603090020146065),  
(3, 0.0734718358370645),  
(4, 0.1892994097121204),  
(5, 0.10838874619051501),  
(6, 0.04024218182927669),  
(7, 0.14673479641335554),  
(8, 0.05923488777590923),  
(9, 0.0967301666813349),  
(10, 0.10259783520851541),  
(11, 0.09464970485606021),  
(12, 0.09037128496931669),  
(13, 0.04499212706658476),  
(14, 0.12824729401064427),  
(15, 0.06282808624375433),  
(16, 0.07894736842105264),  
(17, 0.13977653617040256),  
(18, 0.09493290614465533),  
(19, 0.0000000000000000)]
```

```
In [93]: sorted(list(enumerate(similarity[0])),reverse=True)
```

```
Out[93]: [(4805, 0.0),
(4804, 0.0),
(4803, 0.04499212706658476),
(4802, 0.046829290579084706),
(4801, 0.019252140716412975),
(4800, 0.0),
(4799, 0.052631578947368425),
(4798, 0.04223886030955117),
(4797, 0.0),
(4796, 0.0),
(4795, 0.0),
(4794, 0.0),
(4793, 0.05407380704358751),
(4792, 0.0),
(4791, 0.0),
(4790, 0.0582716546748065),
(4789, 0.060833032924035954),
(4788, 0.0),
(4787, 0.019117977822546817),
(4786, 0.0)]
```

```
In [94]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])
```

```
Out[94]: [(0, 1.0000000000000002),
(1216, 0.28676966733820225),
(2409, 0.26901379342448517),
(3730, 0.2605130246476754),
(507, 0.255608593705383),
(539, 0.25038669783359574),
(582, 0.24511108480187255),
(1204, 0.24455799402225922),
(1194, 0.23179316248638276),
(778, 0.23174488732966075),
(4048, 0.2278389747471728),
(1920, 0.2252817784447915),
(61, 0.22269966704152225),
(2786, 0.21853668936906193),
(172, 0.21239769762143662),
(972, 0.2108663315950723),
(322, 0.2105263157894737),
(2333, 0.20443988269091456),
(3608, 0.20437977982832192),
(3609, 0.2035070436680075)]
```

```
In [95]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]
```

```
Out[95]: [(1216, 0.28676966733820225),
(2409, 0.26901379342448517),
(3730, 0.2605130246476754),
(507, 0.255608593705383),
(539, 0.25038669783359574)]
```

```
In [96]: df.iloc[2409].title
```

```
Out[96]: 'Aliens'
```

final function

```
In [97]: def recommend(movie):  
         movie_index=data[data['title']==movie].index[0]  
         distance= similarity[movie_index]  
         movie_list=sorted(list(enumerate(distance)),reverse=True,key=lambda x: x[1])[1:6]  
  
         for i in movie_list:  
             print(data.iloc[i[0]].title)
```

```
In [98]: recommend ('Avatar')
```

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

```
In [99]: recommend ('Iron Man')
```

Iron Man 3
Iron Man 2
Avengers: Age of Ultron
The Avengers
Captain America: Civil War

```
In [100]: recommend('Spider-Man')
```

Spider-Man 3
Spider-Man 2
The Amazing Spider-Man 2
Arachnophobia
Kick-Ass

```
In [101]: recommend('The Dark Knight')
```

The Dark Knight Rises
Batman Begins
Batman Returns
Batman Forever
Batman

```
In [102]: recommend('X-Men')
```

X2
X-Men: The Last Stand
X-Men: Apocalypse
Iron Man 3
X-Men: First Class

```
In [103]: data[data['title']=='The Dark Knight']
```

```
Out[103]:
```

	movie_id	title	tags
65	155	The Dark Knight	batman rais the stake in hi war on crime. with...

```
In [104]: data[data['title']=='The Dark Knight'].index[0]
```

```
Out[104]: 65
```

```
In [105]: data.iloc[1216].title
```

```
Out[105]: 'Aliens vs Predator: Requiem'
```

```
In [106]: import pickle
```

```
In [107]: data.to_dict()
```

```
36: 91314,  
37: 68728,  
38: 102382,  
39: 20526,  
40: 49013,  
41: 44912,  
42: 10193,  
43: 534,  
44: 168259,  
45: 72190,  
46: 127585,  
47: 54138,  
48: 81005,  
49: 64682,  
50: 9543,  
51: 68726,  
52: 38356,  
53: 217,  
54: 105864,  
55: 62177.
```

```
In [108]: similarity
```

```
Out[108]: 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 [109]: pickle.dump(data.to_dict(),open('movie_dict.pkl',mode='wb'))
```

```
In [110]: pickle.dump(similarity,open('similarity.pkl',mode='wb'))
```