

```
import numpy as np
import pandas as pd
movies=pd.read_csv('tmdb_5000_movies.csv')
credits=pd.read_csv('tmdb_5000_credits.csv')
movies.head(1)
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popular
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...	150.4375

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

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

```
credits=pd.read_csv('tmdb_5000_credits.csv')
credits.head(1)
```

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

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

```
array([{"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": "Andrew Menzies"}, {"credit_id": "549598c3c3a3686ae9004383", "department": "Visual Effects", "gender": 0, "id": 6690, "job": "Visual Effects Producer", "name": "Jill Brooks"}, {"credit_id": "52fe48009251416c750aca4b", "department": "Production", "gender": 1, "id": 6347, "job": "Co-
```

z1e68009251416c750aca4b", "department": "Production", "gender": 1, "id": 6347, "job": "Casting", "name": "Margery Simkin"}, {"credit_id": "570b6f419251417da70032fe", "department": "Art", "gender": 2, "id": 6878, "job": "Supervising Art Director", "name": "Kevin Ishio ka"}, {"credit_id": "5495a0fac3a3686ae9004468", "department": "Sound", "gender": 0, "id": 6883, "job": "Music Editor", "name": "Dick Bernstein"}, {"credit_id": "54959706c3a3686af3003e81", "department": "Sound", "gender": 0, "id": 8159, "job": "Sound Effects Editor", "name": "Shannon Mills"}, {"credit_id": "54491d58c3a3680fb1001ccb", "department": "Sound", "gender": 0, "id": 8160, "job": "Foley", "name": "Dennie Thorpe"}, {"credit_id": "54491d6cc3a3680fa5001b2c", "department": "Sound", "gender": 0, "id": 8163, "job": "Foley", "name": "Jana Vance"}, {"credit_id": "52fe48009251416c750aca57", "department": "Costume & Make-Up", "gender": 1, "id": 8527, "job": "Costume Design", "name": "Deborah Lynn Scott"}, {"credit_id": "52fe48009251416c750aca2f", "department": "Production", "gender": 2, "id": 8529, "job": "Producer", "name": "Jon Landau"}, {"credit_id": "539c4937c3a36810ba002194", "department": "Art", "gender": 0, "id": 9618, "job": "Art Direction", "name": "Sean Haworth"}, {"credit_id": "539c49b6c3a36810c10020e6", "department": "Art", "gender": 1, "id": 12653, "job": "Set Decoration", "name": "Kim Sinclair"}, {"credit_id": "570b6f2f9251413a0e00020d", "department": "Art", "gender": 1, "id": 12653, "job": "Supervising Art Director", "name": "Kim Sinclair"}, {"credit_id": "54491a6c0e0a26748c001b19", "department": "Art", "gender": 2, "id": 14350, "job": "Set Designer", "name": "Richard F. Mays"}, {"credit_id": "56928cf4c3a3684cff0025c4", "department": "Production", "gender": 1, "id": 20294, "job": "Executive Producer", "name": "Laeta Kalogridis"}, {"credit_id": "52fe48009251416c750aca51", "department": "Costume & Make-Up", "gender": 0, "id": 17675, "job": "Costume Design", "name": "Mayes C. Rubeo"}, {"credit_id": "52fe48009251416c750aca11", "department": "Camera", "gender": 2, "id": 18265, "job": "Director of Photography", "name": "Mauro Fiore"}, {"credit_id": "5449194d0e0a26748f001b39", "department": "Art", "gender": 0, "id": 42281, "job": "Set Designer", "name": "Scott Herbertson"}, {"credit_id": "52fe48009251416c750aca05", "department": "Crew", "gender": 0, "id": 42288, "job": "Stunts", "name": "Woody Schultz"}, {"credit_id": "5592aefb92514152de0010f5", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Makeup Artist", "name": "Linda DeVetta"}, {"credit_id": "5592afa92514152de00112c", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Hairstylist", "name": "Linda DeVetta"}, {"credit_id": "54959ed592514130fc002e5d", "department": "Camera", "gender": 2, "id": 33302, "job": "Camera Operator", "name": "Richard Bluck"}, {"credit_id": "539c4891c3a36810ba002147", "department": "Art", "gender": 2, "id": 33303, "job": "Art Direction", "name": "Simon Bright"}, {"credit_id": "54959c069251417a81001f3a", "department": "Visual Effects", "gender": 0, "id": 113145, "job": "Visual Effects Supervisor", "name": "Richard Martin"}, {"credit_id": "54959a0dc3a3680ff5002c8d", "department": "Crew", "gender": 2, "id": 58188, "job": "Visual Effects Editor", "name": "Steve R. Moore"}, {"credit_id": "52fe48009251416c750acald", "department": "Editing", "gender": 2, "id": 58871, "job": "Editor", "name": "John Refoua"}, {"credit_id": "54491a4dc3a3680fc30018ca", "department": "Art", "gender": 0, "id": 92359, "job": "Set Designer", "name": "Karl J. Martin"}, {"credit_id": "52fe48009251416c750aca35", "department": "Camera", "gender": 1, "id": 72201, "job": "Director of Photography", "name": "Chiling Lin"}, {"credit_id": "52fe48009251416c750ac9ff", "department": "Crew", "gender": 0, "id": 89714, "job": "Stunts", "name": "Ilram Choi"}, {"credit_id": "54959c529251416e2b004394", "department": "Visual Effects", "gender": 2, "id": 93214, "job": "Visual Effects Supervisor", "name": "Steven Quale"}, {"credit_id": "54491edf0e0a267489001c37", "department": "Crew", "gender": 1, "id": 122607, "job": "Dialect Coach", "name": "Carla Meyer"}, {"credit_id": "539c485bc3a368653d001a3a", "department": "Art", "gender": 2, "id": 132585, "job": "Art Direction", "name": "Nick Bassett"}, {"credit_id": "539c4903c3a368653d001a74", "department": "Art", "gender": 0, "id": 132596, "job": "Art Direction", "name": "Jill Cormack"}, {"credit_id": "539c4967c3a368653d001a94", "department": "Art", "gender": 0, "id": 132604, "job": "Art Direction", "name": "Andy McLaren"}, {"credit_id": "52fe48009251416c750aca45", "department": "Crew", "gender": 0, "id": 236696, "job": "Motion Capture Artist", "name": "Terry Notary"}, {"credit_id": "54959e02c3a3680fc60027d2", "department": "Crew", "gender": 2, "id": 956198, "job": "Stunt Coordinator", "name": "Garrett Warren"}, {"credit_id": "54959ca3c3a3686ae300438c", "department": "Visual Effects", "gender": 2, "id": 957874, "job": "Visual Effects Supervisor", "name": "Jonathan Rothbart"}, {"credit_id": "570b6f519251412c74001b2f", "department": "Art", "gender": 0, "id": 957889, "job": "Supervising Art Director", "name": "Stefan Dechant"}, {"credit_id": "570b6f62c3a3680b77007460", "department": "Art", "gender": 2, "id": 959555, "job": "Supervising Art Director", "name": "Todd Cherniawsky"}, {"credit_id": "539c4a3ac3a36810da0021cc", "department": "Production", "gender": 0, "id": 1016177, "job": "Casting", "name": "Miranda Rivers"}, {"credit_id": "539c482cc3a36810c1002062", "department": "Art", "gender": 0, "id": 1032536, "job": "Production Design", "name": "Robert Stromberg"}, {"credit_id": "539c4b65c3a36810c9002125", "department": "Costume & Make-Up", "gender": 2, "id": 1071680, "job": "Costume Design", "name": "John Harding"}, {"credit_id": "54959e6692514130fc002e4e", "department": "Camera", "gender": 0, "id": 1177364, "job": "Steadicam Operator", "name": "Roberto De Angelis"}, {"credit_id": "539c49flc3a368653d001aac", "department": "Costume & Make-Up", "gender": 2, "id": 1202850, "job": "Makeup Department Head", "name": "Mike Smithson"}, {"credit_id": "5495999ec3a3686ae100460c", "department": "Visual Effects", "gender": 0, "id": 1204668, "job": "Visual Effects Producer", "name": "Alain Lalanne"}, {"credit_id": "54959cdffc3a3681153002729", "department": "Visual Effects", "gender": 0, "id": 1206410, "job": "Visual Effects Supervisor", "name": "Timothy Galtsoff"}, {"credit_id": "54959c229251417a81001f3a", "department": "Camera", "gender": 2, "id": 1206410, "job": "Camera Operator", "name": "Richard Bluck"}]

Lucas Salton", {"credit_id": "549596239251417a81001eae", "department": "Crew", "gender": 0, "id": 1234266, "job": "Post Production Supervisor", "name": "Janace Tashjian"}, {"credit_id": "54959c859251416e1e003efe", "department": "Visual Effects", "gender": 0, "id": 1271932, "job": "Visual Effects Supervisor", "name": "Stephen Rosenbaum"}, {"credit_id": "5592af28c3a368775a00105f", "department": "Costume & Make-Up", "gender": 0, "id": 1310064, "job": "Makeup Artist", "name": "Frankie Karena"}, {"credit_id": "539c4adfc3a36810e300203b", "department": "Costume & Make-Up", "gender": 1, "id": 1319844, "job": "Costume Supervisor", "name": "Lisa Lovaas"}, {"credit_id": "54959b579251416e2b004371", "department": "Visual Effects", "gender": 0, "id": 1327028, "job": "Visual Effects Supervisor", "name": "Jonathan Fawkner"}, {"credit_id": "539c48a7c3a36810b5001fa7", "department": "Art", "gender": 0, "id": 1330561, "job": "Art Direction", "name": "Robert Bavin"}, {"credit_id": "539c4a71c3a36810da0021e0", "department": "Costume & Make-Up", "gender": 0, "id": 1330567, "job": "Costume Supervisor", "name": "Anthony Almaraz"}, {"credit_id": "539c4a8ac3a36810ba0021e4", "department": "Costume & Make-Up", "gender": 0, "id": 1330570, "job": "Costume Supervisor", "name": "Carolyn M. Fenton"}, {"credit_id": "539c4ab6c3a36810da0021f0", "department": "Costume & Make-Up", "gender": 0, "id": 1330574, "job": "Costume Supervisor", "name": "Beth Koenigsberg"}, {"credit_id": "54491ab70e0a267480001ba2", "department": "Art", "gender": 0, "id": 1336191, "job": "Set Designer", "name": "Sam Page"}, {"credit_id": "544919d9c3a3680fc30018bd", "department": "Art", "gender": 0, "id": 1339441, "job": "Set Designer", "name": "Tex Kadonaga"}, {"credit_id": "54491cf50e0a267483001b0c", "department": "Editing", "gender": 0, "id": 1352422, "job": "Dialogue Editor", "name": "Kim Foscatto"}, {"credit_id": "544919f40e0a26748c001b09", "department": "Art", "gender": 0, "id": 1352962, "job": "Set Designer", "name": "Tammy S. Lee"}, {"credit_id": "5495a115c3a3680ff5002d71", "department": "Crew", "gender": 0, "id": 1357070, "job": "Transportation Coordinator", "name": "Denny Cairra"}, {"credit_id": "5495a12f92514130fc002e94", "department": "Crew", "gender": 0, "id": 1357071, "job": "Transportation Coordinator", "name": "James Waitkus"}, {"credit_id": "5495976fc3a36811530026b0", "department": "Sound", "gender": 0, "id": 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"}, {"credit_id": "54491b79c3a3680fbb001826", "department": "Art", "gender": 0, "id": 1376897, "job": "Construction Coordinator", "name": "John Villarino"}, {"credit_id": "54491baec3a3680fb4001ce6", "department": "Art", "gender": 2, "id": 1376898, "job": "Assistant Art Director", "name": "Jeffrey Wisniewski"}, {"credit_id": "54491d2fc3a3680fb4001d07", "department": "Editing", "gender": 0, "id": 1376899, "job": "Dialogue Editor", "name": "Cheryl Nardi"}, {"credit_id": "54491d86c3a3680fa5001b2f", "department": "Editing", "gender": 0, "id": 1376901, "job": "Dialogue Editor", "name": "Marshall Winn"}, {"credit_id": "54491d9dc3a3680faa001bb0", "department": "Sound", "gender": 0, "id": 1376902, "job": "Supervising Sound Editor", "name": "Gwendolyn Yates Whittle"}, {"credit_id": "54491dc10e0a267486001bce", "department": "Sound", "gender": 0, "id": 1376903, "job": "Sound Re-Recording Mixer", "name": "William Stein"}, {"credit_id": "54491f500e0a26747c001c07", "department": "Crew", "gender": 0, "id": 1376909, "job": "Choreographer", "name": "Lula Washington"}, {"credit_id": "549599239251412c4e002a2e", "department": "Visual Effects", "gender": 0, "id": 1391692, "job": "Visual Effects Producer", "name": "Chris Del Conte"}, {"credit_id": "54959d54c3a36831b8001d9a", "department": "Visual Effects", "gender": 2, "id": 1391695, "job": "Visual Effects Supervisor", "name": "R. Christopher White"}, {"credit_id": "54959bdf9251412c4e002a66", "department": "Visual Effects", "gender": 0, "id": 1394070, "job": "Visual Effects Supervisor", "name": "Dan Lemmon"}, {"credit_id": "5495971d92514132ed002922", "department": "Sound", "gender": 0, "id": 1394129, "job": "Sound Effects Editor", "name": "Tim Nielsen"}, {"credit_id": "5592b25792514152cc0011aa", "department": "Crew", "gender": 0, "id": 1394286, "job": "CG Supervisor", "name": "Michael Mulholland"}, {"credit_id": "54959a329251416e2b004355", "department": "Crew", "gender": 0, "id": 1394750, "job": "Visual Effects Editor", "name": "Thomas Nittmann"}, {"credit_id": "54959d6dc3a3686ae9004401", "department": "Visual Effects", "gender": 0, "id": 1394755, "job": "Visual Effects Supervisor", "name": "Edson Williams"}, {"credit_id": "5495a08fc3a3686ae300441c", "department": "Editing", "gender": 0, "id": 1394953, "job": "Digital Intermediate", "name": "Christine Carr"}, {"credit_id": "55402d659251413d6d000249", "department": "Visual Effects", "gender": 0, "id": 1395269, "job": "Visual Effects Supervisor", "name": "John Bruno"}, {"credit_id": "54959e7b9251416e1e003f3e", "department": "Camera", "gender": 0, "id": 1398970, "job": "Steadicam Operator", "name": "David Emmerichs"}, {"credit_id": "54959734c3a3686ae10045e0", "department": "Sound", "gender": 0, "id": 1400906, "job": "Sound Effects Editor", "name": "Christopher Scarabosio"}, {"credit_id": "54959f3392514130fc002d70", "department": "Crew", "gender": 0, "id": 1400907, "job": "Transportation Coordinator", "name": "Denny Cairra"}]

"credit_id": "549595d2925141301c002d79", "department": "Production", "gender": 0, "id": 1401784, "job": "Production Supervisor", "name": "Jennifer Teves"}, {"credit_id": "549596e892514130fc002d99", "department": "Sound", "gender": 0, "id": 1401785, "job": "Production Manager", "name": "Brigitte Yorke"}, {"credit_id": "549598229251412c4e002alc", "department": "Crew", "gender": 0, "id": 1401786, "job": "Sound Effects Editor", "name": "Ken Fischer"}, {"credit_id": "549598349251416e2b00432b", "department": "Crew", "gender": 0, "id": 1401787, "job": "Special Effects Coordinator", "name": "Iain Hutton"}, {"credit_id": "54959905c3a3686ae3004324", "department": "Visual Effects", "gender": 0, "id": 1401788, "job": "Special Effects Coordinator", "name": "Steve Ingram"}, {"credit_id": "5495994b92514132ed002951", "department": "Visual Effects", "gender": 0, "id": 1401789, "job": "Visual Effects Producer", "name": "Joyce Cox"}, {"credit_id": "549599cbc3a3686ael004613", "department": "Crew", "gender": 0, "id": 1401790, "job": "Visual Effects Producer", "name": "Jenny Foster"}, {"credit_id": "549599f2c3a3686ael00461e", "department": "Crew", "gender": 0, "id": 1401791, "job": "Visual Effects Editor", "name": "Christopher Marino"}, {"credit_id": "54959a51c3a3686af3003eb5", "department": "Visual Effects", "gender": 0, "id": 1401792, "job": "Visual Effects Editor", "name": "Jim Milton"}, {"credit_id": "54959a7cc3a36811530026f4", "department": "Crew", "gender": 0, "id": 1401793, "job": "Visual Effects Producer", "name": "Cyndi Ochs"}, {"credit_id": "54959b91c3a3680ff5002cb4", "department": "Visual Effects", "gender": 0, "id": 1401794, "job": "Visual Effects Editor", "name": "Lucas Putnam"}, {"credit_id": "54959bb69251412c4e002a5f", "department": "Visual Effects", "gender": 0, "id": 1401795, "job": "Visual Effects Supervisor", "name": "Anthony 'Max' Ivins"}, {"credit_id": "54959cbbc3a3686ae3004391", "department": "Visual Effects", "gender": 2, "id": 1401796, "job": "Visual Effects Supervisor", "name": "John Knoll"}, {"credit_id": "54959d06c3a3686ae90043f6", "department": "Visual Effects", "gender": 0, "id": 1401797, "job": "Visual Effects Supervisor", "name": "Eric Saindon"}, {"credit_id": "54959d259251416ele003f11", "department": "Visual Effects", "gender": 0, "id": 1401800, "job": "Visual Effects Supervisor", "name": "Wayne Stables"}, {"credit_id": "54959db49251413af7002975", "department": "Visual Effects", "gender": 0, "id": 1401801, "job": "Visual Effects Supervisor", "name": "David Stinnett"}, {"credit_id": "54959de4c3a3681153002750", "department": "Crew", "gender": 0, "id": 1401803, "job": "Visual Effects Supervisor", "name": "Guy Williams"}, {"credit_id": "54959ef2c3a3680fc60027f2", "department": "Lighting", "gender": 0, "id": 1401804, "job": "Stunt Coordinator", "name": "Stuart Thorp"}, {"credit_id": "54959f07c3a3680fc60027f9", "department": "Camera", "gender": 2, "id": 1401805, "job": "Best Boy Electric", "name": "Giles Coburn"}, {"credit_id": "54959f47c3a3681153002774", "department": "Lighting", "gender": 0, "id": 1401806, "job": "Still Photographer", "name": "Mark Fellman"}, {"credit_id": "54959f8cc3a36831b8001df2", "department": "Visual Effects", "gender": 0, "id": 1401807, "job": "Lighting Technician", "name": "Scott Sprague"}, {"credit_id": "54959fa0c3a36831b8001dfb", "department": "Visual Effects", "gender": 0, "id": 1401808, "job": "Animation Director", "name": "Jeremy Hollobon"}, {"credit_id": "54959fb6c3a3686af3003f54", "department": "Visual Effects", "gender": 0, "id": 1401809, "job": "Animation Director", "name": "Orlando Meunier"}, {"credit_id": "54959fd2c3a36831b8001e02", "department": "Costume & Make-Up", "gender": 0, "id": 1401810, "job": "Animation Director", "name": "Taisuke Tanimura"}, {"credit_id": "54959ff9c3a3686ae300440c", "department": "Costume & Make-Up", "gender": 0, "id": 1401812, "job": "Set Costumer", "name": "Lilia Mishel Acevedo"}, {"credit_id": "5495a0ddc3a3686ael0046fe", "department": "Editing", "gender": 0, "id": 1401814, "job": "Set Costumer", "name": "Alejandro M. Hernandez"}, {"credit_id": "5495a1f7c3a3686ae3004443", "department": "Production", "gender": 0, "id": 1401815, "job": "Digital Intermediate", "name": "Marvin Hall"}, {"credit_id": "5592b29fc3a36869d100002f", "department": "Crew", "gender": 0, "id": 1401816, "job": "Publicist", "name": "Judy Alley"}, {"credit_id": "5592b23a9251415df8001081", "department": "Crew", "gender": 0, "id": 1418381, "job": "CG Supervisor", "name": "Mike Perry"}, {"credit_id": "55491e1192514104c40002d8", "department": "Art", "gender": 0, "id": 1426854, "job": "CG Supervisor", "name": "Andrew Morley"}, {"credit_id": "5525d5809251417276002b06", "department": "Crew", "gender": 0, "id": 1438901, "job": "Conceptual Design", "name": "Seth Engstrom"}, {"credit_id": "554427ca925141586500312a", "department": "Visual Effects", "gender": 0, "id": 1447362, "job": "Visual Effects Art Director", "name": "Eric Oliver"}, {"credit_id": "551906889251415aab001c88", "department": "Art", "gender": 0, "id": 1447503, "job": "Modeling", "name": "Matsune Suzuki"}, {"credit_id": "5592af8492514152cc0010de", "department": "Costume & Make-Up", "gender": 0, "id": 1447524, "job": "Art Department Manager", "name": "Paul Tobin"}, {"credit_id": "553d3c109251415852001318", "department": "Lighting", "gender": 0, "id": 1452643, "job": "Hairstylist", "name": "Roxane Griffin"}, {"credit_id": "5592af4692514152d5001355", "department": "Costume & Make-Up", "gender": 0, "id": 1453938, "job": "Lighting Artist", "name": "Arun Ram-Mohan"}, {"credit_id": "5592b2eac3a36877470012a5", "department": "Crew", "gender": 0, "id": 1457305, "job": "Makeup Artist", "name": "Georgia Lockhart-Adams"}, {"credit_id": "5592b032c3a36877450015f1", "department": "Crew", "gender": 0, "id": 1466035, "job": "CG Supervisor", "name": "Thrain Shadbolt"}, {"credit_id": "5592b0592514152d80012f6", "department": "Crew", "gender": 0, "id": 1483220, "job": "CG Supervisor", "name": "Brad Alexander"}, {"credit_id": "5592b090c3a36877570010b5", "department": "Crew", "gender": 0, "id": 1483221, "job": "CG Supervisor", "name": "Shadi Almassizadeh"}, {"credit_id": "5592b0dbc3a368774b00112c", "department": "Crew", "gender": 0, "id": 1483222, "job": "CG Supervisor", "name": "Simon Clutterbuck"}, {"credit_id": "5592b0592514152d80012f6", "department": "Crew", "gender": 0, "id": 1483223, "job": "CG Supervisor", "name": "Simon Clutterbuck"}]

```
: "Graeme Demmocks"}, {"credit_id": "5592b01e92514152db0010c1", "department": "Crew", "gender": 0, "id": 1483224, "job": "CG Supervisor", "name": "Adrian Fernandes"}, {"credit_id": "5592b11f9251415df8001059", "department": "Crew", "gender": 0, "id": 1483225, "job": "CG Supervisor", "name": "Mitch Gates"}, {"credit_id": "5592b15dc3a3687745001645", "department": "Crew", "gender": 0, "id": 1483226, "job": "CG Supervisor", "name": "Jerry Kung"}, {"credit_id": "5592b18e925141645a0004ae", "department": "Crew", "gender": 0, "id": 1483227, "job": "CG Supervisor", "name": "Andy Lomas"}, {"credit_id": "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"}]]],
dtype=object)
```

In [6]:

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

Out[6]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popular
0	237000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	http://www.avatarmovie.com/	19995	[[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.4375

In [7]:

```
movies.shape
```

Out[7]:

(4803, 20)

In [8]:

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

In [9]:

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

Out[9]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popular
0	237000000	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	http://www.avatarmovie.com/	19995	[[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.4375

1 rows x 23 columns

In [10]:

```
movies['original_language'].value_counts()
```

```

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

```

	genres	id	keywords	original_title	overview	cast	crew
0	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	19995	[{"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...

```
genres          0
id              0
keywords        0
original_title  0
overview        3
cast            0
crew            0
dtype: int64
```

```
In [14]:
movies.dropna(inplace=True)
```

```
In [15]:
movies.duplicated().sum()
```

Out[15]:
0

```
In [16]:
movies.iloc[0].genres
```

Out[16]:
'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'

```
In [17]:
import ast
ast.literal_eval(movies.iloc[0].genres)
```

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

```
In [18]:
def convert(obj):
    L= []
    for i in ast.literal_eval(obj):
        L.append(i['name'])
    return L
```

```
In [19]:
convert(movies.iloc[0].genres)
```

Out[19]:
['Action', 'Adventure', 'Fantasy', 'Science Fiction']

```
In [20]:
movies['genres']=movies['genres'].apply(convert)
```

```
In [21]:
movies.head()
```

Out[21]:

	genres	id	keywords	original_title	overview	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[{"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	[Adventure, Fantasy, Action]	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
	[Action		[{"id": 470,		A cryptic	[{"cast_id": 1,	[{"credit_id":

2	[Action, Adventure, Crime]	206647	["name": "James Bond", "keywords": "spy", {"id": 818, "name": "dc comics"}, {"id": 853, "name": "based on novel"}]	original_title	message from Bond's past	"character": "James Bond", "cast": "54805967c3a36829b5002c41", "credit_id": "54805967c3a36829b5002c41", "crew": "54805967c3a36829b5002c41", "de..."
			818, "name": "dc comics"}, {"id": 853, "name": "based on novel"}]		sends him o...	"cr...
3	[Action, Crime, Drama, Thriller]	49026	["id": 849, "name": "dc comics"}, {"id": 853, "name": "based on novel"}]	The Dark Knight Rises	Following the death of District Attorney Harve...	["cast_id": 2, "character": "Bruce Wayne / Batman", "cast": "52fe4781c3a36847f81398c3", "credit_id": "52fe4781c3a36847f81398c3", "crew": "52fe4781c3a36847f81398c3", "de..."]
4	[Action, Adventure, Science Fiction]	49529	["id": 818, "name": "based on novel"}, {"id": "..."]	John Carter	John Carter is a war-weary, former military ca...	["cast_id": 5, "character": "John Carter", "cast": "52fe479ac3a36847f813eaa3", "credit_id": "52fe479ac3a36847f813eaa3", "crew": "52fe479ac3a36847f813eaa3", "de..."]

In [22]:

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

In [23]:

```
movies.head()
```

Out[23]:

	genres	id	keywords	original_title	overview	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[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	[Adventure, Fantasy, Action]	285	[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 Spa...	["credit_id": "52fe4232c3a36847f800b579", "de...
2	[Action, Adventure, Crime]	206647	[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", "cr...	["credit_id": "54805967c3a36829b5002c41", "de...
3	[Action, Crime, Drama, Thriller]	49026	[dc comics, crime fighter, terrorist, secret i...	The Dark Knight Rises	Following the death of District Attorney Harve...	["cast_id": 2, "character": "Bruce Wayne / Ba...	["credit_id": "52fe4781c3a36847f81398c3", "de...
4	[Action, Adventure, Science Fiction]	49529	[based on novel, mars, medallion, space travel...	John Carter	John Carter is a war-weary, former military ca...	["cast_id": 5, "character": "John Carter", "c...	["credit_id": "52fe479ac3a36847f813eaa3", "de...

In [24]:

```
def convert3(obj):
    L= []
    counter=0;
    for i in ast.literal_eval(obj):
        if counter != 3:
            L.append(i['name'])
            counter+=1
        else:
            break
    return L
```

In [25]:

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

In [26]:

```
movies.head()
```

Out[26]:

	genres	id	keywords	original_title	overview	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[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	[Adventure, Fantasy, Action]	285	[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]	["credit_id": "52fe4232c3a36847f800b579", "de...
2	[Action, Adventure, Crime]	206647	[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	[Action, Crime, Drama, Thriller]	49026	[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	[Action, Adventure, Science Fiction]	49529	[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 [27]:

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

In [29]:

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

In [30]:

```
movies.head()
```

Out[30]:

	genres	id	keywords	original_title	overview	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[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	[Adventure, Fantasy, Action]	285	[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	[Action, Adventure, Crime]	206647	[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	[Action, Crime, Drama, Thriller]	49026	[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	[Action, Adventure, Science Fiction]	49529	[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 [31]:

```
movies['overview'][0]
```

Out[31]:

Out[31]:

'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 [32]:

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

In [33]:

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

Out[33]:

	genres	id	keywords	original_title	overview	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[culture clash, future, space war, space colon...	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]

In [34]:

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

In [35]:

```
movies.head()
```

Out[35]:

	genres	id	keywords	original_title	overview	cast	crew	tags
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[culture clash, future, space war, space colon...	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]	[In, the, 22nd, century,, a, paraplegic, Marin...
1	[Adventure, Fantasy, Action]	285	[ocean, drug abuse, exotic island, east india ...	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]	[Captain, Barbossa,, long, believed, to, be, d...
2	[Action, Adventure, Crime]	206647	[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]	[A, cryptic, message, from, Bond's, past, send...
3	[Action, Crime, Drama, Thriller]	49026	[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]	[Following, the, death, of, District, Attorney...
4	[Action, Adventure, Science Fiction]	49529	[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]	[John, Carter, is, a, war-weary,, former, mili...

In [36]:

```
movies['genres'] = movies['genres'].apply(lambda x:[i.replace(" ", "") for i in x])
```

In [37]:

```
movies['keywords'] = movies['keywords'].apply(lambda x:[i.replace(" ", "") for i in x])
movies['cast'] = movies['cast'].apply(lambda x:[i.replace(" ", "") for i in x])
movies['crew'] = movies['crew'].apply(lambda x:[i.replace(" ", "") for i in x])
```

In [38]:

```
movies.head()
```

Out[38]:

	genres	id	keywords	original_title	overview	cast	crew	tags
0	[Action, Adventure, Fantasy, ScienceFiction]	19995	[cultureclash, future, spacewar, spacecolony, ...	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	[In, the, 22nd, century,, a, paraplegic, Marin...
1	[Adventure, Fantasy, Action]	285	[ocean, drugabuse, exoticisland, eastindiatrad...	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	[Action, Adventure, Crime]	206647	[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	[Action, Crime, Drama, Thriller]	49026	[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	[Action, Adventure, ScienceFiction]	49529	[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 [39]:

```
new_df=movies[['id','original_title','tags']]
```

In [40]:

```
new_df.head()
```

Out[40]:

	id	original_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 [50]:

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

Out[50]:

'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. a ction adventure fantasy science fiction culture clash future space war space colony society space travel futuristic romance space alien tribe alien planet cgi marine soldier battle love affair anti war power relations mind and soul 3d sam worthington zoe saldana sigourney weaverjames cameron'

In [42]:

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

C:\Users\Nandani\AppData\Local\Temp\ipykernel_12856\1824047427.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 [46]:

```
new_df.head()
```

Out[46]:

	id	original_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 [48]:

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

C:\Users\Nandani\AppData\Local\Temp\ipykernel_12856\3923394162.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
new_df['tags']=new_df['tags'].apply(lambda x : x.lower())

In [49]:

```
new_df.head()
```

Out[49]:

	id	original_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 [51]:

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

Out[51]:

```
'in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique m  
ission, but becomes torn between following orders and protecting an alien civilization. a  
ction adventure fantasy science fiction culture clash future space war space colony socie  
ty space travel futuristic romance space alien tribe alien planet cgi marine soldier batt  
le love affair anti war power relations mind and soul 3d sam worthington zoe saldana sig  
urney weaverjames cameron'
```

In [53]:

```
movies.head()
```

Out[53]:

	genres	id	keywords	original_title	overview	cast	crew	tags
0	[Action, Adventure, Fantasy, ScienceFiction]	19995	[cultureclash, future, spacewar, spacecolony, ...]	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	[In, the, 22nd, century,, a, paraplegic, Marin...
1	[Adventure, Fantasy, Action]	285	[ocean, drugabuse, exoticisland, eastindiatrad...]	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	[Action, Adventure, Crime]	206647	[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	[Action, Crime, Drama, Thriller]	49026	[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	[Action, Adventure, ScienceFiction]	49529	[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 [76]:

```
from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000,stop_words='english')
```

In [71]:

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

In [72]:

```
vectors
```

Out[72]:

```
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 [73]:

```
vectors[2]
```

Out[73]:

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

In [74]:

```
cv.get_feature_names()
```

Out[74]:

0ae[74].

['000',
'007',
'10',
'100',
'11',
'12',
'13',
'14',
'15',
'16',
'16th',
'17',
'17th',
'18',
'18th',
'19',
'1910',
'1920',
'1930',
'1940',
'1950',
'1950s',
'1960',
'1960s',
'1970',
'1970s',
'1980',
'1985',
'1990',
'19th',
'20',
'200',
'2009',
'20th',
'21st',
'24',
'25',
'30',
'300',
'3d',
'40',
'50',
'500',
'51',
'60',
'70',
'aaron',
'abandon',
'abbi',
'abduct',
'abel',
'abigail',
'abil',
'abl',
'aboard',
'aborigin',
'abov',
'abraham',
'abroad',
'abus',
'academi',
'accent',
'accept',
'access',
'accid',
'accident',
'acclaim',
'accompani',
'accomplish',
'account',
'accus',

'ace',
'achiev',
'act',
'action',
'activ',
'activist',
'activities',
'actor',
'actress',
'actual',
'ad',
'adam',
'adapt',
'add',
'addict',
'adjust',
'admir',
'admit',
'adolesc',
'adolf',
'adopt',
'ador',
'adrian',
'adrien',
'adult',
'adulteri',
'adulthood',
'advanc',
'adventur',
'adventure',
'advertis',
'advic',
'advis',
'affair',
'affect',
'affleck',
'afghanistan',
'africa',
'african',
'aftercreditssting',
'afterlif',
'aftermath',
'ag',
'age',
'agenc',
'agency',
'agenda',
'agent',
'agents',
'aggress',
'ago',
'agre',
'ahead',
'aid',
'aidan',
'ail',
'aim',
'air',
'aircraft',
'airplan',
'airport',
'aka',
'al',
'alabama',
'alan',
'alaska',
'alba',
'albert',
'alcohol',
'alec',
'alejandro',
'alex',

'alexand',
'alexi',
'alfr',
'ali',
'alic',
'alice',
'alicia',
'alien',
'alik',
'alison',
'aliv',
'alive',
'allan',
'allen',
'alli',
'allianc',
'allison',
'allow',
'alon',
'alongsid',
'alpha',
'alreadi',
'alter',
'altern',
'always',
'alyssa',
'alzheimer',
'amanda',
'amateur',
'amaz',
'amazon',
'ambassador',
'amber',
'ambigu',
'ambit',
'ambiti',
'ambul',
'ambush',
'america',
'american',
'ami',
'amid',
'amidst',
'amnesia',
'amp',
'amsterdam',
'amus',
'amy',
'ana',
'analyst',
'anarch',
'ancient',
'anderson',
'andi',
'andrea',
'andrew',
'android',
'andré',
'andy',
'ang',
'angarano',
'angel',
'angela',
'angeles',
'angelina',
'anger',
'angri',
'angst',
'ani',
'anim',
'animals',
'aniston',

'ann',
'anna',
'annabel',
'annett',
'anni',
'annie',
'anniversari',
'announc',
'annual',
'anonym',
'anoth',
'answer',
'ant',
'antholog',
'anthoni',
'anthropomorph',
'anti',
'antoin',
'anton',
'antonio',
'anyon',
'anyth',
'apart',
'apartheid',
'apartment',
'ape',
'apocalyps',
'apocalypt',
'appar',
'appear',
'appl',
'appleg',
'appoint',
'appreci',
'apprentic',
'approach',
'april',
'arab',
'arch',
'archaeologist',
'archeolog',
'archer',
'architect',
'arctic',
'area',
'aren',
'arena',
'argument',
'aris',
'aristocrat',
'arkin',
'arm',
'armand',
'armi',
'armor',
'armstrong',
'army',
'arnold',
'aronofski',
'arquett',
'arrang',
'arrest',
'arriv',
'arrog',
'arrow',
'art',
'arthur',
'articl',
'artifact',
'artifici',
'artist',
'ash',

'ashley',
'ashton',
'asia',
'asian',
'asid',
'ask',
'aspect',
'aspir',
'assassin',
'assault',
'assign',
'assist',
'assistant',
'associ',
'assum',
'asteroid',
'astin',
'astronaut',
'asylum',
'athlet',
'atkin',
'atlant',
'atom',
'attack',
'attacks',
'attempt',
'attend',
'attent',
'attitud',
'attorney',
'attract',
'auction',
'audienc',
'audit',
'audrey',
'august',
'aunt',
'austin',
'australia',
'australian',
'austria',
'author',
'autism',
'auto',
'autobiograph',
'automobil',
'aveng',
'averag',
'avoid',
'await',
'awaken',
'award',
'away',
'awkward',
'awri',
'awry',
'ax',
'axe',
'aykroyd',
'babe',
'babi',
'baby',
'bachchan',
'bachelor',
'backdrop',
'background',
'backpack',
'bacon',
'bad',
'bag',
'bahama',
'bail',

'bailey',
'baker',
'balanc',
'baldwin',
'bale',
'ball',
'ballet',
'balloon',
'baltimor',
'ban',
'bana',
'band',
'bandera',
'bandit',
'bangkok',
'banish',
'bank',
'banker',
'banner',
'bar',
'barbara',
'barbarian',
'bardem',
'bare',
'bargain',
'barn',
'barney',
'baron',
'barri',
'barrier',
'barrymor',
'bart',
'bas',
'base',
'basebal',
'basement',
'basing',
'basketbal',
'basketball',
'bassett',
'bat',
'bate',
'bateman',
'bath',
'bathroom',
'batman',
'battl',
'battle',
'battlefield',
'bay',
'beach',
'beam',
'bean',
'bear',
'beard',
'beast',
'beat',
'beatti',
'beauti',
'beautiful',
'beauty',
'becam',
'becaus',
'beck',
'becker',
'becki',
'beckinsal',
'becom',
'bed',
'bedroom',
'bee',
'beer',

'befor',
'befriend',
'began',
'begin',
'begins',
'behavior',
'belgium',
'belief',
'believ',
'bell',
'bella',
'bello',
'belong',
'belov',
'belt',
'ben',
'bend',
'bene',
'beneath',
'benefit',
'benicio',
'benjamin',
'bennett',
'bent',
'bentley',
'berg',
'bergman',
'berlin',
'bernal',
'berni',
'berri',
'best',
'bet',
'beth',
'betray',
'bett',
'bettani',
'better',
'betti',
'beverli',
'bibl',
'biblic',
'bicycl',
'biel',
'big',
'bigg',
'bigger',
'biggest',
'bike',
'biker',
'bikini',
'billi',
'billion',
'billionair',
'biographi',
'biolog',
'bird',
'birth',
'birthday',
'bisexu',
'bishop',
'bit',
'bite',
'bitten',
'bitter',
'bizarr',
'black',
'blackmail',
'blacksmith',
'blade',
'blair',
'blake',

'blame',
'blanchett',
'bledel',
'blend',
'bless',
'blind',
'bliss',
'block',
'blond',
'blood',
'bloodi',
'bloodthirsti',
'bloom',
'blow',
'blue',
'blunt',
'board',
'boat',
'bob',
'bobb',
'bodi',
'body',
'bodyguard',
'bold',
'bollywood',
'bomb',
'bond',
'bone',
'bonham',
'bonni',
'book',
'booth',
'border',
'bore',
'boredom',
'born',
'boss',
'boston',
'bosworth',
'botch',
'bottl',
'bound',
'boundari',
'bounti',
'bower',
'bowl',
'bowman',
'box',
'boxer',
'boy',
'boyd',
'boyfriend',
'boyl',
'boys',
'brad',
'bradford',
'bradley',
'brain',
'branagh',
'brand',
'brando',
'brandon',
'bratt',
'brave',
'braveri',
'brazil',
'brazilian',
'break',
'breakdown',
'breast',
'breath',
'breed',

'brenda',
'brendan',
'brent',
'breslin',
'brett',
'brian',
'bride',
'bridesmaid',
'bridg',
'bridget',
'brie',
'brief',
'brien',
'bright',
'brilliant',
'bring',
'brink',
'britain',
'british',
'britt',
'brittani',
'broadbent',
'broadcast',
'broadway',
'broderick',
'brodi',
'broke',
'broken',
'broker',
'brolin',
'bronx',
'brook',
'brooklyn',
'broom',
'brosnan',
'brothel',
'brother',
'brothers',
'brought',
'brown',
'bruce',
'bruno',
'brutal',
'bryan',
'bryce',
'bu',
'buck',
'bud',
'buddi',
'buddy',
'budget',
'build',
'building',
'built',
'bullet',
'bulli',
'bullock',
'bumbl',
'bunni',
'burglar',
'buri',
'burk',
'burn',
'burstyn',
'burt',
'burton',
'buscemi',
'bush',
'busi',
'business',
'businessman',
'bust',

'butcher',
'butler',
'butt',
'buy',
'buzz',
'byrn',
'caan',
'cabin',
'cabl',
'caesar',
'cage',
'cain',
'cairo',
'cal',
'california',
'calvin',
'camcord',
'came',
'camera',
'cameraman',
'cameron',
'camp',
'campaign',
'campbel',
'campbell',
'campu',
'canada',
'canadian',
'cancer',
'candi',
'candid',
'canin',
'cannib',
'cannon',
'capabl',
'caper',
'capit',
'capt',
'captain',
'captiv',
'captur',
'car',
'card',
'care',
'career',
'carefre',
'carel',
'caretak',
'carey',
'cargo',
'cari',
'caribbean',
'carl',
'carla',
'carlo',
'carmen',
'carol',
'carolin',
'carolina',
'carpent',
'carradin',
'carrey',
'carri',
'carrie',
'carrier',
'carrol',
'cartel',
'carter',
'cartoon',
'case',
'casey',
'cash',

'casino',
'cassidi',
'cast',
'castl',
'castle',
'casual',
'cat',
'cataclysm',
'catastroph',
'catch',
'cate',
'catherin',
'cathol',
'catholic',
'cattl',
'caught',
'caus',
'cavalri',
'cave',
'caviezel',
'cedric',
'celebr',
'cell',
'cemeteri',
'center',
'centr',
'central',
'centuri',
'centuries',
'century',
'certain',
'chad',
'chain',
'chainsaw',
'chair',
'challeng',
'chamber',
'champion',
'championship',
'chan',
'chanc',
'chance',
'chang',
'change',
'changes',
'channel',
'chao',
'chaos',
'chaotic',
'chapter',
'charact',
'character',
'characters',
'charg',
'charismat',
'charl',
'charli',
'charlie',
'charliz',
'charlott',
'charm',
'chart',
'chase',
'chastain',
'chat',
'chauffeur',
'chazz',
'theadl',
'cheat',
'check',
'cheerlead',
'chef',

'chelsea',
'chemic',
'chen',
'cher',
'cheryl',
'chestnut',
'chevi',
'chicago',
'chicken',
'chief',
'child',
'childhood',
'children',
'chill',
'china',
'chines',
'chip',
'chipmunk',
'chiwetel',
'chloe',
'chloë',
'chocol',
'choic',
'choice',
'choir',
'choos',
'chopra',
'chosen',
'chow',
'chri',
'christ',
'christensen',
'christian',
'christin',
'christina',
'christma',
'christmas',
'christoph',
'chronicl',
'chuck',
'church',
'cia',
'cigar',
'cigarett',
'cillian',
'cinema',
'cinemat',
'circl',
'circu',
'circuit',
'circumst',
'citi',
'citizen',
'city',
'civil',
'civilian',
'claim',
'clair',
'clan',
'clark',
'clash',
'class',
'classic',
'classmat',
'classroom',
'clau',
'claud',
'claudia',
'clay',
'clean',
'clear',
'clerk',

'client',
'cliff',
'climat',
'climb',
'clint',
'clive',
'clock',
'clone',
'clooney',
'close',
'closer',
'closet',
'cloud',
'clown',
'club',
'clue',
'clumsi',
'clutch',
'coach',
'coal',
'coast',
'cocain',
'code',
'codi',
'coen',
'coffin',
'cohen',
'col',
'cold',
'cole',
'coleman',
'colin',
'collaps',
'colleagu',
'collect',
'collector',
'colleg',
'college',
'collett',
'collid',
'collin',
'collis',
'colonel',
'coloni',
'color',
'colorado',
'columbu',
'coma',
'combat',
'combin',
'come',
'comeback',
'comed',
'comedi',
'comedian',
'comedy',
'comet',
'comfort',
'comic',
'coming',
'command',
'commando',
'commerci',
'commit',
'common',
'commun',
'community',
'compani',
'companion',
'company',
'compet',
'competit',

'competition',
'complet',
'complex',
'complic',
'compos',
'compuls',
'comput',
'conan',
'concentr',
'concern',
'concert',
'condit',
'condition',
'confeder',
'confer',
'confess',
'confid',
'confin',
'conflict',
'confront',
'confus',
'congress',
'connect',
'connel',
'connelli',
'conneri',
'conni',
'connor',
'conquer',
'conrad',
'consequ',
'consequences',
'conserv',
'consid',
'conspir',
'conspiraci',
'conspiracy',
'constant',
'constantli',
'construct',
'consum',
'contact',
'contain',
'contemporari',
'contend',
'content',
'contest',
'continu',
'contract',
'contractor',
'control',
'controversi',
'conveni',
'convent',
'converg',
'convers',
'convert',
'convict',
'convinc',
'cook',
'cool',
'cooper',
'cop',
'cope',
'coppola',
...]

In [59]:

```
import nltk  
from nltk.stem.porter import PorterStemmer
```

```
ps=PorterStemmer()
```

In [64]:

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

In [63]:

```
ps.stem('loved')
```

Out[63]:

```
'love'
```

In [65]:

```
stem('in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization. action adventure fantasy science fiction culture clash future space war space colony society space travel futuristic romance space alien tribe alien planet cgi marine soldier battle love affair anti war power relations mind and soul 3d sam worthington zoe saldana sigourney weaver james cameron')
```

Out[65]:

```
'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a unique mission , but becom torn between follow order and protect an alien civilization. action adventur fantasi scienc fiction cultur clash futur space war space coloni societi space travel futurist romanc space alien tribe alien planet cgi marin soldier battl love affair anti war power relat mind and soul 3d sam worthington zoe saldana sigourney weaver jame cameron'
```

In [67]:

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

C:\Users\Nandani\AppData\Local\Temp\ipykernel_12856\3514595201.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

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

In [79]:

```
import sklearn  
from sklearn.metrics.pairwise import cosine_similarity
```

In [83]:

```
similarity = cosine_similarity(vectors)
```

In [86]:

```
similarity[2]
```

Out[86]:

```
array([0.06554735, 0.04167571, 1.          , ..., 0.01844278, 0.0808122 ,  
       0.          ])
```

In [87]:

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

Out[87]:

```
[(2409, 0.4935676473963735),
```

```
(4336, 0.4113063728303113),  
(1537, 0.3880712182748949),  
(3162, 0.3879852850993631),  
(373, 0.3804429551263411)]
```

In [100]:

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

In [102]:

```
recommend('Batman Begins')
```

```
The Dark Knight  
The Dark Knight Rises  
Batman  
Batman  
Batman & Robin
```

In [96]:

```
new_df.iloc[1216].original_title
```

Out[96]:

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

In []: