In [1]:

```python
import pandas as pd
import numpy as np
```

In [2]:

```python
movies = pd.read_csv('tmdb_5000_movies.csv')
```

In [3]:

```python
movies.head()
```

Out[3]:

genres	homepage	id	keywords	original_language	original_title	overview	popularity	produ
[{"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...	150.437577	[{" F
[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[{"na Pic
[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond’s past sends him o...	107.376788	[{'
[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"r Pictu
[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[{"na

```
In [4]: credits= pd.read_csv('tmdb_5000_credits.csv')
```

```
In [5]: credits.head()
```

Out[5]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [6]: credits.head(1)['cast'].values
```

```
Out[6]: array(['[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "nam
        e": "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": "52fe4800
        9251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quari
        tch", "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, "nam
        e": "Giovanni Ribisi", "order": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id": "52fe48009251416c750ac9dd", "ge
        nder": 2, "id": 59231, "name": "Joel David Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_id": "52fe4800925
        1416c750ac9e5", "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": "T
        su\'Tey", "credit_id": "52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso", "order": 9}, {"cast_id": 1
        2, "character": "Dr. Max Patel", "credit_id": "52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "name": "Dileep Rao", "or
        der": 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", "g
        ender": 2, "id": 154153, "name": "Sean Anthony Moran", "order": 12}, {"cast_id": 33, "character": "Cryo Vault Med Tech", "cre
        dit_id": "52fe48009251416c750aca5f", "gender": 2, "id": 397312, "name": "Jason Whyte", "order": 13}, {"cast_id": 34, "charact
        er": "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": 986
        734, "name": "Kelly Kilgour", "order": 15}, {"cast_id": 36, "character": "Shuttle Pilot", "credit_id": "52fe48009251416c750ac
        a6b", "gender": 0, "id": 1207227, "name": "James Patrick Pitt", "order": 16}, {"cast_id": 37, "character": "Shuttle Co-Pilo
        t", "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": "Pet
        er 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, "charac
        ter": "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": "52fe48009251416c750
        aca87", "gender": 0, "id": 143206, "name": "Michael Blain-Rozgay", "order": 23}, {"cast_id": 44, "character": "Suit #2", "cre
        dit_id": "52fe48009251416c750aca8b", "gender": 2, "id": 169676, "name": "Jon Curry", "order": 24}, {"cast_id": 46, "characte
        r": "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": 4
        2288, "name": "Woody Schultz", "order": 26}, {"cast_id": 48, "character": "Horse Clan Leader", "credit_id": "52fe48009251416c
        750aca97", "gender": 2, "id": 68278, "name": "Peter Mensah", "order": 27}, {"cast_id": 49, "character": "Link Room Tech", "cr
        edit_id": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247, "name": "Sonia Yee", "order": 28}, {"cast_id": 50, "charact
        er": "Basketball Avatar / Troupe", "credit_id": "52fe48009251416c750aca9f", "gender": 1, "id": 1207248, "name": "Jahnel Curfm
        an", "order": 29}, {"cast_id": 51, "character": "Basketball Avatar", "credit_id": "52fe48009251416c750acaa3", "gender": 0, "i
        d": 89714, "name": "Ilram Choi", "order": 30}, {"cast_id": 52, "character": "Na\'vi Child", "credit_id": "52fe48009251416c750
        acaa7", "gender": 0, "id": 1207249, "name": "Kyla Warren", "order": 31}, {"cast_id": 53, "character": "Troupe", "credit_id":
```

"52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Roumain", "order": 32}, {"cast_id": 54, "character": "T
roupe", "credit_id": "52fe48009251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id":
57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "orde
r": 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, "i
d": 1207252, "name": "Jodie Landau", "order": 36}, {"cast_id": 58, "character": "Troupe", "credit_id": "52fe48009251416c750ac
abf", "gender": 0, "id": 1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "character": "Troupe", "credit_id": "52
fe48009251416c750acac3", "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": 115
8600, "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": "52fe4801
9251416c750acad7", "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": "N
athan Meister", "order": 45}, {"cast_id": 67, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae3", "gende
r": 0, "id": 1207264, "name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op Center Staff", "credit_id": "52fe
48019251416c750acae7", "gender": 2, "id": 33311, "name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "character": "O
p Center Staff", "credit_id": "52fe48019251416c750acaeb", "gender": 0, "id": 1207265, "name": "Paul Yates", "order": 48}, {"c
ast_id": 70, "character": "Op Center Duty Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 1207266, "nam
e": "Wray Wilson", "order": 49}, {"cast_id": 71, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaf3", "ge
nder": 2, "id": 54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "character": "Dancer", "credit_id": "52fe4801925
1416c750acaf7", "gender": 0, "id": 1207267, "name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 73, "character": "Danc
er", "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", "or
der": 53}, {"cast_id": 75, "character": "Dancer", "credit_id": "52fe48019251416c750acb03", "gender": 0, "id": 1207270, "nam
e": "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": "52fe4801925141
6c750acb0b", "gender": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "c
redit_id": "52fe48019251416c750acb0f", "gender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_id": 79, "c
haracter": "Dancer", "credit_id": "52fe48019251416c750acb13", "gender": 0, "id": 1207274, "name": "April Marie Thomas", "orde
r": 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": "52fe48019251416c7
50acb1b", "gender": 0, "id": 1207276, "name": "Colin Bleasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (u
ncredited)", "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": "M
att Clayton", "order": 62}, {"cast_id": 84, "character": "Nav\'i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gen
der": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_i
d": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "characte
r": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "orde

```
r": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2,
"id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "c
redit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 14, "c
haracter": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "orde
r": 68}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 120
1399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 90, "character": "Banshee (uncredited)", "credit_id": "52fe480192
51416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 91, "character": "Soldier (uncr
edited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_i
d": 92, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "n
ame": "Logan Pithyou", "order": 72}, {"cast_id": 93, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c7
50acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 94, "character": "Hero (uncredite
d)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 95, "chara
cter": "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 (uncredite
d)", "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": "Jodi
e Taylor", "order": 78}, {"cast_id": 99, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb6
3", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 100, "character": "Geologist (uncredi
ted)", "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": "N
ikie Zambo", "order": 81}, {"cast_id": 102, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6
f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}]'],
      dtype=object)
```

```
In [7]: credits.head(1)['crew'].values
```

```
Out[7]: 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": "Producti
        on Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900,
        "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gend
        er": 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": "5544ee3b925141499
        f0008fc", "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 Camer
        on"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "gender": 2, "id": 2710, "job": "Writer", "name": "J
        ames Cameron"}, {"credit_id": "52fe48009251416c750aca17", "department": "Editing", "gender": 2, "id": 2710, "job": "Editor",
        "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca29", "department": "Production", "gender": 2, "id": 2710, "jo
        b": "Producer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca3f", "department": "Writing", "gender": 2, "i
        d": 2710, "job": "Screenplay", "name": "James Cameron"}, {"credit_id": "539c4987c3a36810ba0021a4", "department": "Art", "gend
        er": 2, "id": 7236, "job": "Art Direction", "name": "Andrew Menzies"}, {"credit_id": "549598c3c3a3686ae9004383", "departmen
        t": "Visual Effects", "gender": 0, "id": 6690, "job": "Visual Effects Producer", "name": "Jill Brooks"}, {"credit_id": "52fe4
        8009251416c750aca4b", "department": "Production", "gender": 1, "id": 6347, "job": "Casting", "name": "Margery Simkin"}, {"cre
        dit_id": "570b6f419251417da70032fe", "department": "Art", "gender": 2, "id": 6878, "job": "Supervising Art Director", "name":
        "Kevin Ishioka"}, {"credit_id": "5495a0fac3a3686ae9004468", "department": "Sound", "gender": 0, "id": 6883, "job": "Music Edi
        tor", "name": "Dick Bernstein"}, {"credit_id": "54959706c3a3686af3003e81", "department": "Sound", "gender": 0, "id": 8159, "j
        ob": "Sound Effects Editor", "name": "Shannon Mills"}, {"credit_id": "54491d58c3a3680fb1001ccb", "department": "Sound", "gend
        er": 0, "id": 8160, "job": "Foley", "name": "Dennie Thorpe"}, {"credit_id": "54491d6cc3a3680fa5001b2c", "department": "Soun
        d", "gender": 0, "id": 8163, "job": "Foley", "name": "Jana Vance"}, {"credit_id": "52fe48009251416c750aca57", "department":
        "Costume & Make-Up", "gender": 1, "id": 8527, "job": "Costume Design", "name": "Deborah Lynn Scott"}, {"credit_id": "52fe4800
        9251416c750aca2f", "department": "Production", "gender": 2, "id": 8529, "job": "Producer", "name": "Jon Landau"}, {"credit_i
        d": "539c4937c3a36810ba002194", "department": "Art", "gender": 0, "id": 9618, "job": "Art Direction", "name": "Sean Hawort
        h"}, {"credit_id": "539c49b6c3a36810c10020e6", "department": "Art", "gender": 1, "id": 12653, "job": "Set Decoration", "nam
        e": "Kim Sinclair"}, {"credit_id": "570b6f2f9251413a0e00020d", "department": "Art", "gender": 1, "id": 12653, "job": "Supervi
        sing Art Director", "name": "Kim Sinclair"}, {"credit_id": "54491a6c0e0a26748c001b19", "department": "Art", "gender": 2, "i
        d": 14350, "job": "Set Designer", "name": "Richard F. Mays"}, {"credit_id": "56928cf4c3a3684cff0025c4", "department": "Produc
        tion", "gender": 1, "id": 20294, "job": "Executive Producer", "name": "Laeta Kalogridis"}, {"credit_id": "52fe48009251416c750
        aca51", "department": "Costume & Make-Up", "gender": 0, "id": 17675, "job": "Costume Design", "name": "Mayes C. Rubeo"}, {"cr
        edit_id": "52fe48009251416c750aca11", "department": "Camera", "gender": 2, "id": 18265, "job": "Director of Photography", "na
        me": "Mauro Fiore"}, {"credit_id": "5449194d0e0a26748f001b39", "department": "Art", "gender": 0, "id": 42281, "job": "Set Des
        igner", "name": "Scott Herbertson"}, {"credit_id": "52fe48009251416c750aca05", "department": "Crew", "gender": 0, "id": 4228
        8, "job": "Stunts", "name": "Woody Schultz"}, {"credit_id": "5592aefb92514152de0010f5", "department": "Costume & Make-Up", "g
        ender": 0, "id": 29067, "job": "Makeup Artist", "name": "Linda DeVetta"}, {"credit_id": "5592afa492514152de00112c", "departme
        nt": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Hairstylist", "name": "Linda DeVetta"}, {"credit_id": "54959ed592
        514130fc002e5d", "department": "Camera", "gender": 2, "id": 33302, "job": "Camera Operator", "name": "Richard Bluck"}, {"cred
        it_id": "539c4891c3a36810ba002147", "department": "Art", "gender": 2, "id": 33303, "job": "Art Direction", "name": "Simon Bri
        ght"}, {"credit_id": "54959c069251417a81001f3a", "department": "Visual Effects", "gender": 0, "id": 113145, "job": "Visual Ef
```

```
fects Supervisor", "name": "Richard Martin"}, {"credit_id": "54959a0dc3a3680ff5002c8d", "department": "Crew", "gender": 2, "i
d": 58188, "job": "Visual Effects Editor", "name": "Steve R. Moore"}, {"credit_id": "52fe48009251416c750aca1d", "department":
"Editing", "gender": 2, "id": 58871, "job": "Editor", "name": "John Refoua"}, {"credit_id": "54491a4dc3a3680fc30018ca", "depa
rtment": "Art", "gender": 0, "id": 92359, "job": "Set Designer", "name": "Karl J. Martin"}, {"credit_id": "52fe48009251416c75
0aca35", "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"}, {"c
redit_id": "54959c529251416e2b004394", "department": "Visual Effects", "gender": 2, "id": 93214, "job": "Visual Effects Super
visor", "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", "depart
ment": "Art", "gender": 0, "id": 132604, "job": "Art Direction", "name": "Andy McLaren"}, {"credit_id": "52fe48009251416c750a
ca45", "department": "Crew", "gender": 0, "id": 236696, "job": "Motion Capture Artist", "name": "Terry Notary"}, {"credit_i
d": "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", "dep
artment": "Art", "gender": 2, "id": 959555, "job": "Supervising Art Director", "name": "Todd Cherniawsky"}, {"credit_id": "53
9c4a3ac3a36810da0021cc", "department": "Production", "gender": 0, "id": 1016177, "job": "Casting", "name": "Miranda Rivers"},
{"credit_id": "539c482cc3a36810c1002062", "department": "Art", "gender": 0, "id": 1032536, "job": "Production Design", "nam
e": "Robert Stromberg"}, {"credit_id": "539c4b65c3a36810c9002125", "department": "Costume & Make-Up", "gender": 2, "id": 1071
680, "job": "Costume Design", "name": "John Harding"}, {"credit_id": "54959e6692514130fc002e4e", "department": "Camera", "gen
der": 0, "id": 1177364, "job": "Steadicam Operator", "name": "Roberto De Angelis"}, {"credit_id": "539c49f1c3a368653d001aac",
"department": "Costume & Make-Up", "gender": 2, "id": 1202850, "job": "Makeup Department Head", "name": "Mike Smithson"}, {"c
redit_id": "5495999ec3a3686ae100460c", "department": "Visual Effects", "gender": 0, "id": 1204668, "job": "Visual Effects Pro
ducer", "name": "Alain Lalanne"}, {"credit_id": "54959cdfc3a3681153002729", "department": "Visual Effects", "gender": 0, "i
d": 1206410, "job": "Visual Effects Supervisor", "name": "Lucas Salton"}, {"credit_id": "549596239251417a81001eae", "departme
nt": "Crew", "gender": 0, "id": 1234266, "job": "Post Production Supervisor", "name": "Janace Tashjian"}, {"credit_id": "5495
9c859251416e1e003efe", "department": "Visual Effects", "gender": 0, "id": 1271932, "job": "Visual Effects Supervisor", "nam
e": "Stephen Rosenbaum"}, {"credit_id": "5592af28c3a368775a00105f", "department": "Costume & Make-Up", "gender": 0, "id": 131
0064, "job": "Makeup Artist", "name": "Frankie Karena"}, {"credit_id": "539c4adfc3a36810e300203b", "department": "Costume & M
ake-Up", "gender": 1, "id": 1319844, "job": "Costume Supervisor", "name": "Lisa Lovaas"}, {"credit_id": "54959b579251416e2b00
4371", "department": "Visual Effects", "gender": 0, "id": 1327028, "job": "Visual Effects Supervisor", "name": "Jonathan Fawk
ner"}, {"credit_id": "539c48a7c3a36810b5001fa7", "department": "Art", "gender": 0, "id": 1330561, "job": "Art Direction", "na
me": "Robert Bavin"}, {"credit_id": "539c4a71c3a36810da0021e0", "department": "Costume & Make-Up", "gender": 0, "id": 133056
7, "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": "539c4ab6c3a
36810da0021f0", "department": "Costume & Make-Up", "gender": 0, "id": 1330574, "job": "Costume Supervisor", "name": "Beth Koe
nigsberg"}, {"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 D
esigner", "name": "Tex Kadonaga"}, {"credit_id": "54491cf50e0a267483001b0c", "department": "Editing", "gender": 0, "id": 1352
```

```
422, "job": "Dialogue Editor", "name": "Kim Foscato"}, {"credit_id": "544919f40e0a26748c001b09", "department": "Art", "gende
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": "5495a12f92514
130fc002e94", "department": "Crew", "gender": 0, "id": 1357071, "job": "Transportation Coordinator", "name": "James Waitku
s"}, {"credit_id": "5495976fc3a36811530026b0", "department": "Sound", "gender": 0, "id": 1360103, "job": "Supervising Sound E
ditor", "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", "departme
nt": "Art", "gender": 0, "id": 1376890, "job": "Set Designer", "name": "Jonathan Dyer"}, {"credit_id": "54491967c3a3680faa001
b5e", "department": "Art", "gender": 0, "id": 1376891, "job": "Set Designer", "name": "Joseph Hiura"}, {"credit_id": "5449199
7c3a3680fb1001c8a", "department": "Art", "gender": 0, "id": 1376892, "job": "Art Department Coordinator", "name": "Rebecca Je
llie"}, {"credit_id": "544919ba0e0a26748f001b42", "department": "Art", "gender": 0, "id": 1376893, "job": "Set Designer", "na
me": "Robert Andrew Johnson"}, {"credit_id": "54491b1dc3a3680faa001b8c", "department": "Art", "gender": 0, "id": 1376895, "jo
b": "Assistant Art Director", "name": "Mike Stassi"}, {"credit_id": "54491b79c3a3680fbb001826", "department": "Art", "gende
r": 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_i
d": "54491d2fc3a3680fb4001d07", "department": "Editing", "gender": 0, "id": 1376899, "job": "Dialogue Editor", "name": "Chery
l Nardi"}, {"credit_id": "54491d86c3a3680fa5001b2f", "department": "Editing", "gender": 0, "id": 1376901, "job": "Dialogue Ed
itor", "name": "Marshall Winn"}, {"credit_id": "54491d9dc3a3680faa001bb0", "department": "Sound", "gender": 0, "id": 1376902,
"job": "Supervising Sound Editor", "name": "Gwendolyn Yates Whittle"}, {"credit_id": "54491dc10e0a267486001bce", "departmen
t": "Sound", "gender": 0, "id": 1376903, "job": "Sound Re-Recording Mixer", "name": "William Stein"}, {"credit_id": "54491f50
0e0a26747c001c07", "department": "Crew", "gender": 0, "id": 1376909, "job": "Choreographer", "name": "Lula Washington"}, {"cr
edit_id": "549599239251412c4e002a2e", "department": "Visual Effects", "gender": 0, "id": 1391692, "job": "Visual Effects Prod
ucer", "name": "Chris Del Conte"}, {"credit_id": "54959d54c3a36831b8001d9a", "department": "Visual Effects", "gender": 2, "i
d": 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"}, {"cred
it_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 Superv
isor", "name": "Michael Mulholland"}, {"credit_id": "54959a329251416e2b004355", "department": "Crew", "gender": 0, "id": 1394
750, "job": "Visual Effects Editor", "name": "Thomas Nittmann"}, {"credit_id": "54959d6dc3a3686ae9004401", "department": "Vis
ual Effects", "gender": 0, "id": 1394755, "job": "Visual Effects Supervisor", "name": "Edson Williams"}, {"credit_id": "5495a
08fc3a3686ae300441c", "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, "i
d": 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": "549595d
d92514130fc002d79", "department": "Production", "gender": 0, "id": 1401784, "job": "Production Supervisor", "name": "Jennifer
Teves"}, {"credit_id": "549596009251413af70028cc", "department": "Production", "gender": 0, "id": 1401785, "job": "Production
Manager", "name": "Brigitte Yorke"}, {"credit_id": "549596e892514130fc002d99", "department": "Sound", "gender": 0, "id": 1401
786, "job": "Sound Effects Editor", "name": "Ken Fischer"}, {"credit_id": "549598229251412c4e002a1c", "department": "Crew",
"gender": 0, "id": 1401787, "job": "Special Effects Coordinator", "name": "Iain Hutton"}, {"credit_id": "549598349251416e2b00
432b", "department": "Crew", "gender": 0, "id": 1401788, "job": "Special Effects Coordinator", "name": "Steve Ingram"}, {"cre
dit_id": "54959905c3a3686ae3004324", "department": "Visual Effects", "gender": 0, "id": 1401789, "job": "Visual Effects Produ
cer", "name": "Joyce Cox"}, {"credit_id": "5495994b92514132ed002951", "department": "Visual Effects", "gender": 0, "id": 1401
790, "job": "Visual Effects Producer", "name": "Jenny Foster"}, {"credit_id": "549599cbc3a3686ae1004613", "department": "Cre
w", "gender": 0, "id": 1401791, "job": "Visual Effects Editor", "name": "Christopher Marino"}, {"credit_id": "549599f2c3a3686
ae100461e", "department": "Crew", "gender": 0, "id": 1401792, "job": "Visual Effects Editor", "name": "Jim Milton"}, {"credit
_id": "54959a51c3a3686af3003eb5", "department": "Visual Effects", "gender": 0, "id": 1401793, "job": "Visual Effects Produce
r", "name": "Cyndi Ochs"}, {"credit_id": "54959a7cc3a36811530026f4", "department": "Crew", "gender": 0, "id": 1401794, "job":
"Visual Effects Editor", "name": "Lucas Putnam"}, {"credit_id": "54959b91c3a3680ff5002cb4", "department": "Visual Effects",
"gender": 0, "id": 1401795, "job": "Visual Effects Supervisor", "name": "Anthony \'Max\' Ivins"}, {"credit_id": "54959bb69251
412c4e002a5f", "department": "Visual Effects", "gender": 0, "id": 1401796, "job": "Visual Effects Supervisor", "name": "John
Knoll"}, {"credit_id": "54959cbbc3a3686ae3004391", "department": "Visual Effects", "gender": 2, "id": 1401799, "job": "Visual
Effects Supervisor", "name": "Eric Saindon"}, {"credit_id": "54959d06c3a3686ae90043f6", "department": "Visual Effects", "gend
er": 0, "id": 1401800, "job": "Visual Effects Supervisor", "name": "Wayne Stables"}, {"credit_id": "54959d259251416e1e003f1
1", "department": "Visual Effects", "gender": 0, "id": 1401801, "job": "Visual Effects Supervisor", "name": "David Stinnet
t"}, {"credit_id": "54959db49251413af7002975", "department": "Visual Effects", "gender": 0, "id": 1401803, "job": "Visual Eff
ects Supervisor", "name": "Guy Williams"}, {"credit_id": "54959de4c3a3681153002750", "department": "Crew", "gender": 0, "id":
1401804, "job": "Stunt Coordinator", "name": "Stuart Thorp"}, {"credit_id": "54959ef2c3a3680fc60027f2", "department": "Lighti
ng", "gender": 0, "id": 1401805, "job": "Best Boy Electric", "name": "Giles Coburn"}, {"credit_id": "54959f07c3a3680fc60027f
9", "department": "Camera", "gender": 2, "id": 1401806, "job": "Still Photographer", "name": "Mark Fellman"}, {"credit_id":
"54959f47c3a3681153002774", "department": "Lighting", "gender": 0, "id": 1401807, "job": "Lighting Technician", "name": "Scot
t Sprague"}, {"credit_id": "54959f8cc3a36831b8001df2", "department": "Visual Effects", "gender": 0, "id": 1401808, "job": "An
imation Director", "name": "Jeremy Hollobon"}, {"credit_id": "54959fa0c3a36831b8001dfb", "department": "Visual Effects", "gen
der": 0, "id": 1401809, "job": "Animation Director", "name": "Orlando Meunier"}, {"credit_id": "54959fb6c3a3686af3003f54", "d
epartment": "Visual Effects", "gender": 0, "id": 1401810, "job": "Animation Director", "name": "Taisuke Tanimura"}, {"credit_
id": "54959fd2c3a36831b8001e02", "department": "Costume & Make-Up", "gender": 0, "id": 1401812, "job": "Set Costumer", "nam
e": "Lilia Mishel Acevedo"}, {"credit_id": "54959ff9c3a3686ae300440c", "department": "Costume & Make-Up", "gender": 0, "id":
1401814, "job": "Set Costumer", "name": "Alejandro M. Hernandez"}, {"credit_id": "5495a0ddc3a3686ae10046fe", "department": "E
diting", "gender": 0, "id": 1401815, "job": "Digital Intermediate", "name": "Marvin Hall"}, {"credit_id": "5495a1f7c3a3686ae3
004443", "department": "Production", "gender": 0, "id": 1401816, "job": "Publicist", "name": "Judy Alley"}, {"credit_id": "55
92b29fc3a36869d100002f", "department": "Crew", "gender": 0, "id": 1418381, "job": "CG Supervisor", "name": "Mike Perry"}, {"c
redit_id": "5592b23a9251415df8001081", "department": "Crew", "gender": 0, "id": 1426854, "job": "CG Supervisor", "name": "And
rew Morley"}, {"credit_id": "55491e1192514104c40002d8", "department": "Art", "gender": 0, "id": 1438901, "job": "Conceptual D
esign", "name": "Seth Engstrom"}, {"credit_id": "5525d5809251417276002b06", "department": "Crew", "gender": 0, "id": 1447362,
"job": "Visual Effects Art Director", "name": "Eric Oliver"}, {"credit_id": "554427ca925141586500312a", "department": "Visual
Effects", "gender": 0, "id": 1447503, "job": "Modeling", "name": "Matsune Suzuki"}, {"credit_id": "551906889251415aab001c88",
"department": "Art", "gender": 0, "id": 1447524, "job": "Art Department Manager", "name": "Paul Tobin"}, {"credit_id": "5592a
f8492514152cc0010de", "department": "Costume & Make-Up", "gender": 0, "id": 1452643, "job": "Hairstylist", "name": "Roxane Gr
iffin"}, {"credit_id": "553d3c109251415852001318", "department": "Lighting", "gender": 0, "id": 1453938, "job": "Lighting Art
ist", "name": "Arun Ram-Mohan"}, {"credit_id": "5592af4692514152d5001355", "department": "Costume & Make-Up", "gender": 0, "i
```

```
d": 1457305, "job": "Makeup Artist", "name": "Georgia Lockhart-Adams"}, {"credit_id": "5592b2eac3a36877470012a5", "departmen
t": "Crew", "gender": 0, "id": 1466035, "job": "CG Supervisor", "name": "Thrain Shadbolt"}, {"credit_id": "5592b032c3a3687745
0015f1", "department": "Crew", "gender": 0, "id": 1483220, "job": "CG Supervisor", "name": "Brad Alexander"}, {"credit_id":
"5592b05592514152d80012f6", "department": "Crew", "gender": 0, "id": 1483221, "job": "CG Supervisor", "name": "Shadi Almassiz
adeh"}, {"credit_id": "5592b090c3a36877570010b5", "department": "Crew", "gender": 0, "id": 1483222, "job": "CG Supervisor",
"name": "Simon Clutterbuck"}, {"credit_id": "5592b0dbc3a368774b00112c", "department": "Crew", "gender": 0, "id": 1483223, "jo
b": "CG Supervisor", "name": "Graeme Demmocks"}, {"credit_id": "5592b0fe92514152db0010c1", "department": "Crew", "gender": 0,
"id": 1483224, "job": "CG Supervisor", "name": "Adrian Fernandes"}, {"credit_id": "5592b11f9251415df8001059", "department":
"Crew", "gender": 0, "id": 1483225, "job": "CG Supervisor", "name": "Mitch Gates"}, {"credit_id": "5592b15dc3a3687745001645",
"department": "Crew", "gender": 0, "id": 1483226, "job": "CG Supervisor", "name": "Jerry Kung"}, {"credit_id": "5592b18e92514
1645a0004ae", "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 Mari
no"}, {"credit_id": "5592b2049251415df8001078", "department": "Crew", "gender": 0, "id": 1483229, "job": "CG Supervisor", "na
me": "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": "5592b37c
c3a368775100113b", "department": "Crew", "gender": 0, "id": 1483234, "job": "CG Supervisor", "name": "Ben White"}, {"credit_i
d": "573c8e2f9251413f5d000094", "department": "Crew", "gender": 1, "id": 1621932, "job": "Stunts", "name": "Min Windle"}]'],
      dtype=object)
```

In [8]:
```python
movies = movies.merge(credits,on='title')
```

In [9]:
```python
movies.head(1)
```

Out[9]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production
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...	150.437577	[{"name" Film Pa

1 rows × 23 columns

In [10]:
```python
movies.shape
```

```
Out[10]:  (4809, 23)
```

```
In [11]:  movies['original_language'].value_counts()
```

Out[11]:

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

In [12]:

```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4809 entries, 0 to 4808
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   budget                4809 non-null   int64  
 1   genres                4809 non-null   object 
 2   homepage              1713 non-null   object 
 3   id                    4809 non-null   int64  
 4   keywords              4809 non-null   object 
 5   original_language     4809 non-null   object 
 6   original_title        4809 non-null   object 
 7   overview              4806 non-null   object 
 8   popularity            4809 non-null   float64
 9   production_companies  4809 non-null   object 
 10  production_countries  4809 non-null   object 
 11  release_date          4808 non-null   object 
 12  revenue               4809 non-null   int64  
 13  runtime               4807 non-null   float64
 14  spoken_languages      4809 non-null   object 
 15  status                4809 non-null   object 
 16  tagline               3965 non-null   object 
 17  title                 4809 non-null   object 
 18  vote_average          4809 non-null   float64
 19  vote_count            4809 non-null   int64  
 20  movie_id              4809 non-null   int64  
 21  cast                  4809 non-null   object 
 22  crew                  4809 non-null   object 
dtypes: float64(3), int64(5), object(15)
memory usage: 864.2+ KB
```

In [13]:

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

In [14]:

```
movies.head()
```

Out[14]:

	movie_id	genres	title	cast	crew	overview	keywords
0	19995	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...	In the 22nd century, a paraplegic Marine is di...	[{"id": 1463, "name": "culture clash"}, {"id":...
1	285	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...	Captain Barbossa, long believed to be dead, ha...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...
2	206647	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...	A cryptic message from Bond's past sends him o...	[{"id": 470, "name": "spy"}, {"id": 818, "name...
3	49026	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...	Following the death of District Attorney Harve...	[{"id": 849, "name": "dc comics"}, {"id": 853,...
4	49529	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...	John Carter is a war-weary, former military ca...	[{"id": 818, "name": "based on novel"}, {"id":...

In [15]:

```
movies.isnull().sum()
```

Out[15]:

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

In [16]:

```
movies.dropna(inplace= True)
```

In [17]:
```
movies.duplicated().sum()
```

Out[17]:
```
np.int64(0)
```

In [18]:
```
movies.iloc[0].genres
```

Out[18]:
```
'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science
Fiction"}]'
```

In [19]:
```
def convert(obj):
    L = []
    for i in ast.literal_eval(obj):
        L.append(i['name'])
    return L
```

In [20]:
```
import ast
ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "
```

Out[20]:
```
[{'id': 28, 'name': 'Action'},
 {'id': 12, 'name': 'Adventure'},
 {'id': 14, 'name': 'Fantasy'},
 {'id': 878, 'name': 'Science Fiction'}]
```

In [21]:
```
 movies['genres'] =  movies['genres'].apply(convert)
```

In [22]:
```
import ast

def convert(obj):
    if isinstance(obj, str):
        # If string, parse it
        obj = ast.literal_eval(obj)
    # Now, obj should be a list of dicts or list of strings
    if isinstance(obj, list) and all(isinstance(i, dict) and 'name' in i for i in obj):
        return [i['name'] for i in obj]
    elif isinstance(obj, list):
        return obj
    return []
```

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

Out[23]:

	movie_id	genres	title	cast	crew	overview	keywords
0	19995	[Action, Adventure, Fantasy, Science Fiction]	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...	In the 22nd century, a paraplegic Marine is di...	[{"id": 1463, "name": "culture clash"}, {"id":...
1	285	[Adventure, Fantasy, Action]	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...	Captain Barbossa, long believed to be dead, ha...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...
2	206647	[Action, Adventure, Crime]	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...	A cryptic message from Bond's past sends him o...	[{"id": 470, "name": "spy"}, {"id": 818, "name...
3	49026	[Action, Crime, Drama, Thriller]	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...	Following the death of District Attorney Harve...	[{"id": 849, "name": "dc comics"}, {"id": 853,...
4	49529	[Action, Adventure, Science Fiction]	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...	John Carter is a war-weary, former military ca...	[{"id": 818, "name": "based on novel"}, {"id":...

```
In [24]:  movies['keywords'] = movies['keywords'].apply(convert)
```

```
In [25]: movies.head()
```

Out[25]:

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

In [26]:

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

    return L
```

In [27]:
```
movies['cast'] = movies['cast'].apply(convert3)
```

In [28]:
```
movies.head()
```

Out[28]:

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

In [29]:
```
def fetch_director(obj):
    L = []
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            L.append(i['name'])
            break
    return L
```

In [30]:
```
movies['crew'] = movies['crew'].apply(fetch_director)
```

In [31]:
```
movies.head()
```

Out[31]:

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

In [32]:
```
movies['overview'][0]
```

Out[32]:
```
'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between fol
lowing orders and protecting an alien civilization.'
```

In [33]:
```
movies['overview'] = movies['overview'].apply(lambda x: x.split())
```

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

Out[34]:

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

In [35]:

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

In [36]:

```
movies.head()
```

Out[36]:

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

In [37]:

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

In [38]:

```
movies.head()
```

Out[38]:

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

In [39]:

```python
new_df = movies[['movie_id','title','tags']]
```

In [40]:

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

```
C:\Users\dell\AppData\Local\Temp\ipykernel_20896\3089450492.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

In [41]:

```python
new_df.head()
```

Out[41]:

	movie_id	title	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	A cryptic message from Bond’s past sends him o...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...
4	49529	John Carter	John Carter is a war-weary, former military ca...

In [42]:

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

Out[42]:

```
'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between fol
lowing orders and protecting an alien civilization. Action Adventure Fantasy ScienceFiction cultureclash future spacewar spac
ecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar pow
errelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCameron'
```

In [43]:

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

```
C:\Users\dell\AppData\Local\Temp\ipykernel_20896\1380776331.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-ve
rsus-a-copy
  new_df['tags'] = new_df['tags'].apply(lambda x: x.lower())
```

In [44]:

```
new_df.head()
```

Out[44]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a paraplegic marine is di...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believed to be dead, ha...
2	206647	Spectre	a cryptic message from bond’s past sends him o...
3	49026	The Dark Knight Rises	following the death of district attorney harve...
4	49529	John Carter	john carter is a war-weary, former military ca...

In [45]:

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

Out[45]:

```
'in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but becomes torn between fol
lowing orders and protecting an alien civilization. action adventure fantasy sciencefiction cultureclash future spacewar spac
ecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar pow
errelations mindandsoul 3d samworthington zoesaldana sigourneyweaver jamescameron'
```

In [46]:

```
new_df['tags'][1]
```

Out[46]:

```
"captain barbossa, long believed to be dead, has come back to life and is headed to the edge of the earth with will turner an
d elizabeth swann. but nothing is quite as it seems. adventure fantasy action ocean drugabuse exoticisland eastindiatradingco
mpany loveofone'slife traitor shipwreck strongwoman ship alliance calypso afterlife fighter pirate swashbuckler aftercreditss
tinger johnnydepp orlandobloom keiraknightley goreverbinski"
```

In [47]:

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

In [48]:

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

Out[48]:

```
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]])
```

In [49]:

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

```
Out[49]:  (4806, 5000)
```

```
In [50]:  vectors = cv.fit_transform(new_df['tags']).toarray()
```

```
In [51]:  vectors
```

```
Out[51]:  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]])
```

```
In [52]:  vectors[0]
```

```
Out[52]:  array([0, 0, 0, ..., 0, 0, 0])
```

```
In [53]:  cv.get_feature_names_out()
```

```
Out[53]:  array(['000', '007', '10', ..., 'zone', 'zoo', 'zooeydeschanel'],
                dtype=object)
```

```
In [54]:  len(cv.get_feature_names_out())
```

```
Out[54]:  5000
```

```
In [55]:  !pip install nltk
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: nltk in c:\users\dell\appdata\roaming\python\python312\site-packages (3.9.1)
Requirement already satisfied: click in c:\users\dell\appdata\roaming\python\python312\site-packages (from nltk) (8.2.1)
Requirement already satisfied: joblib in c:\users\dell\appdata\roaming\python\python312\site-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in c:\users\dell\appdata\roaming\python\python312\site-packages (from nltk) (202
4.11.6)
Requirement already satisfied: tqdm in c:\users\dell\appdata\roaming\python\python312\site-packages (from nltk) (4.67.1)
Requirement already satisfied: colorama in c:\users\dell\appdata\roaming\python\python312\site-packages (from click->nltk) (0.
4.6)
```

```
[notice] A new release of pip is available: 25.0.1 -> 25.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [56]:

```python
import nltk
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
```

In [57]:

```python
def stem(text):
    y = []

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

    return " ".join(y)
```

In [58]:

```python
['loved','loving','love'
,'love','love','love']
```

Out[58]:

```
['loved', 'loving', 'love', 'love', 'love', 'love']
```

In [59]:

```python
ps.stem('loving')
```

Out[59]:

```
'love'
```

In [60]:

```python
ps.stem('danced')
```

Out[60]:

```
'danc'
```

In [61]:

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

```
C:\Users\dell\AppData\Local\Temp\ipykernel_20896\3213734980.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

In [62]:

```
cv.get_feature_names_out()
```

Out[62]:

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

In [63]:

```
from sklearn.feature_extraction.text import CountVectorizer

cv = CountVectorizer(
    stop_words='english',
    lowercase=True,
    token_pattern=r'(?u)\b[a-zA-Z]{3,}\b'  # only words with 3+ English letters
)

X = cv.fit_transform(new_df['tags'])
features = cv.get_feature_names_out()

# Print filtered English feature names
print(features)
```

```
['aaa' 'aaliyah' 'aam' ... 'zurich' 'zuzu' 'zyklon']
```

from sklearn.metrics.pairwise import cosine_similarity

In [64]:

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

In [65]:

```
cosine_similarity(vectors)
```

```
Out[65]: array([[1.        , 0.08740748, 0.05827165, ..., 0.02418254, 0.02564946,
                 0.        ],
                [0.08740748, 1.        , 0.06451613, ..., 0.02677398, 0.        ,
                 0.        ],
                [0.05827165, 0.06451613, 1.        , ..., 0.02677398, 0.        ,
                 0.        ],
                ...,
                [0.02418254, 0.02677398, 0.02677398, ..., 1.        , 0.07071068,
                 0.04836508],
                [0.02564946, 0.        , 0.        , ..., 0.07071068, 1.        ,
                 0.05129892],
                [0.        , 0.        , 0.        , ..., 0.04836508, 0.05129892,
                 1.        ]])
```

```
In [66]: similarity = cosine_similarity(vectors)
```

```
In [67]: similarity
```

```
Out[67]: array([[1.        , 0.08740748, 0.05827165, ..., 0.02418254, 0.02564946,
                 0.        ],
                [0.08740748, 1.        , 0.06451613, ..., 0.02677398, 0.        ,
                 0.        ],
                [0.05827165, 0.06451613, 1.        , ..., 0.02677398, 0.        ,
                 0.        ],
                ...,
                [0.02418254, 0.02677398, 0.02677398, ..., 1.        , 0.07071068,
                 0.04836508],
                [0.02564946, 0.        , 0.        , ..., 0.07071068, 1.        ,
                 0.05129892],
                [0.        , 0.        , 0.        , ..., 0.04836508, 0.05129892,
                 1.        ]])
```

```
In [68]: similarity[0]
```

```
Out[68]: array([1.        , 0.08740748, 0.05827165, ..., 0.02418254, 0.02564946,
                0.        ])
```

```
In [69]: sorted(similarity[0])[-10:-1]
```

```
Out[69]: [np.float64(0.23084512921915967),
          np.float64(0.23084512921915967),
          np.float64(0.23174488732966075),
          np.float64(0.233785950059759),
          np.float64(0.23939494881986934),
          np.float64(0.2409900932515112),
          np.float64(0.24283093212859141),
          np.float64(0.24779731389167606),
          np.float64(0.25038669783359574)]
```

```
In [70]: sorted(similarity[0], reverse = True)
```

```
Out[70]: [np.float64(1.0000000000000002),
          np.float64(0.25038669783359574),
          np.float64(0.24779731389167606),
          np.float64(0.24283093212859141),
          np.float64(0.2409900932515112),
          np.float64(0.23939494881986934),
          np.float64(0.233785950059759),
          np.float64(0.23174488732966075),
          np.float64(0.23084512921915967),
          np.float64(0.23084512921915967),
          np.float64(0.2294157338705618),
          np.float64(0.21677749238103003),
          np.float64(0.21629522817435),
          np.float64(0.21526419295572297),
          np.float64(0.21486752129677),
          np.float64(0.21296183592613546),
          np.float64(0.21239769762143662),
          np.float64(0.21239769762143662),
          np.float64(0.2108663315950723),
          np.float64(0.20857039859669468),
          np.float64(0.20770324619863198),
          np.float64(0.20751433915982237),
          np.float64(0.20395079136182276),
          np.float64(0.20395079136182276),
          np.float64(0.2029530274475215),
          np.float64(0.2029530274475215),
          np.float64(0.2024645717996314),
          np.float64(0.19767387315371682),
          np.float64(0.19466570535691505),
          np.float64(0.19194297398747862),
          np.float64(0.19117977822546817),
          np.float64(0.19088542889273336),
          np.float64(0.19088542889273336),
          np.float64(0.18848425873126295),
          np.float64(0.18731716231633883),
          np.float64(0.18731716231633883),
          np.float64(0.18541926977182405),
          np.float64(0.1853959098637286),
          np.float64(0.1842105263157895),
          np.float64(0.18172434016970185),
          np.float64(0.1813690625275029),
          np.float64(0.1810898182317451),
          np.float64(0.18074256993863339),
          np.float64(0.18074256993863339),
          np.float64(0.17954621161490197),
          np.float64(0.17954621161490197),
          np.float64(0.177343107178349),
          np.float64(0.177343107178349),
          np.float64(0.17699808135119718),
          np.float64(0.17699808135119718),
          np.float64(0.17521916101261562),
          np.float64(0.1745025152241333),
          np.float64(0.17206180040292132),
          np.float64(0.16943474841747155),
          np.float64(0.1692777916923361),
          np.float64(0.1691275228729346),
          np.float64(0.16742770563222897),
          np.float64(0.16556654463313053),
          np.float64(0.16556654463313053),
          np.float64(0.16481712868606585),
          np.float64(0.16452254913212452),
          np.float64(0.16452254913212452),
          np.float64(0.16350382386265752),
          np.float64(0.16222142113076254),
          np.float64(0.16222142113076254),
          np.float64(0.16222142113076254),
          np.float64(0.16222142113076252),
          np.float64(0.16222142113076252),
          np.float64(0.15907119074394446),
          np.float64(0.15907119074394446),
          np.float64(0.15789473684210528),
          np.float64(0.15747244473304667),
          np.float64(0.15609763526361567),
          np.float64(0.15609763526361567),
          np.float64(0.15585730003983933),
          np.float64(0.15585730003983933),
          np.float64(0.15389675281277312),
          np.float64(0.15328483487124145),
          np.float64(0.152008377581442),
          np.float64(0.15061880828219448),
          np.float64(0.15018785229652765),
          np.float64(0.14910501299480672),
          np.float64(0.14808721943977307),
          np.float64(0.14808721943977307),
          np.float64(0.14673479641335554),
          np.float64(0.14673479641335554),
          np.float64(0.14567913668701626),
          np.float64(0.14567913668701626),
          np.float64(0.14509525002200235),
          np.float64(0.14509525002200233),
          np.float64(0.14509525002200233),
          np.float64(0.1442149876003076),
          np.float64(0.14350946197048198),
          np.float64(0.1433848336691011),
          np.float64(0.14159846508095775),
          np.float64(0.14159846508095775),
          np.float64(0.14159846508095775),
          np.float64(0.1411956236812263),
          np.float64(0.1404878717372541),
          np.float64(0.13977653617040256),
          np.float64(0.1391037210186643),
          np.float64(0.1391037210186643),
          np.float64(0.1391037210186643),
          np.float64(0.13834289277321493),
          np.float64(0.13834289277321493),
          np.float64(0.13834289277321493),
          np.float64(0.13764944032233709),
          np.float64(0.13710212427677043),
          np.float64(0.13710212427677043),
          np.float64(0.13530201829834768),
          np.float64(0.13530201829834768),
          np.float64(0.13530201829834768),
          np.float64(0.13518451760896877),
          np.float64(0.13518451760896877),
          np.float64(0.13497638119975428),
          np.float64(0.13497638119975428),
          np.float64(0.13369695534647594),
          np.float64(0.13369695534647594),
          np.float64(0.13334518676566626),
          np.float64(0.1324532357065044),
          np.float64(0.1324532357065044),
          np.float64(0.1324532357065044),
          np.float64(0.13157894736842107),
          np.float64(0.13157894736842107),
          np.float64(0.13157894736842107),
          np.float64(0.13157894736842107),
          np.float64(0.13157894736842107),
          np.float64(0.13112201362143713),
          np.float64(0.1302565123238377),
          np.float64(0.130066495428618),
          np.float64(0.12988108336653278),
          np.float64(0.12988108336653278),
          np.float64(0.12977713690461004),
          np.float64(0.12977713690461004),
          np.float64(0.12977713690461004),
          np.float64(0.12977713690461004),
          np.float64(0.12892051277806202),
          np.float64(0.12824729401064427),
          np.float64(0.12725695259515557),
          np.float64(0.12725695259515557),
          np.float64(0.12725695259515557),
          np.float64(0.12725695259515557),
          np.float64(0.12725695259515557),
          np.float64(0.126673647984535),
          np.float64(0.126673647984535),
          np.float64(0.12565617248750865),
          np.float64(0.12565617248750865),
          np.float64(0.12515654358043973),
          np.float64(0.12515654358043973),
          np.float64(0.12487810821089254),
          np.float64(0.12487810821089254),
          np.float64(0.12369267399882337),
          np.float64(0.12369267399882337),
          np.float64(0.12369267399882337),
          np.float64(0.12262786789699316),
          np.float64(0.12262786789699316),
          np.float64(0.12262786789699313),
          np.float64(0.12262786789699313),
          np.float64(0.12227899701112963),
          np.float64(0.12227899701112963),
          np.float64(0.12227899701112963),
          np.float64(0.12227899701112963),
          np.float64(0.12227899701112963),
          np.float64(0.12166606584807191),
          np.float64(0.12091270835166862),
          np.float64(0.12091270835166862),
          np.float64(0.12091270835166862),
          np.float64(0.1204950466257556),
          np.float64(0.1204950466257556),
          np.float64(0.1204950466257556),
          np.float64(0.11980845957463077),
          np.float64(0.11959121830873498),
          np.float64(0.11959121830873498),
          np.float64(0.11846977555181847),
          np.float64(0.11846977555181847),
          np.float64(0.11846977555181847),
          np.float64(0.11846977555181847),
          np.float64(0.11831213107007527),
          np.float64(0.11831213107007527),
          np.float64(0.11831213107007527),
          np.float64(0.11803342130469505),
          np.float64(0.11803342130469505),
          np.float64(0.11803342130469505),
          np.float64(0.11803342130469505),
          np.float64(0.116543309349613),
          np.float64(0.116543309349613),
          np.float64(0.11587244366483038),
          np.float64(0.11587244366483038),
          np.float64(0.11587244366483038),
          np.float64(0.1147078669352809),
          np.float64(0.1147078669352809),
          np.float64(0.11470786693528087),
          np.float64(0.11470786693528087),
          np.float64(0.11470786693528087),
          np.float64(0.11470786693528087),
          np.float64(0.11470786693528087),
          np.float64(0.1139194873735864),
          np.float64(0.11357771260606365),
          np.float64(0.11357771260606365),
          np.float64(0.11357771260606365),
          np.float64(0.11295649894498103),
          np.float64(0.11295649894498103),
          np.float64(0.11295649894498103),
          np.float64(0.11295649894498103),
          np.float64(0.11295649894498103),
          np.float64(0.11295649894498103),
          np.float64(0.11239029738980326),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11164843913471803),
          np.float64(0.11128297681493143),
          np.float64(0.11128297681493143),
          np.float64(0.11128297681493143),
          np.float64(0.11128297681493143),
          np.float64(0.11037769642208699),
          np.float64(0.10968169942141635),
          np.float64(0.10968169942141635),
          np.float64(0.10968169942141635),
          np.float64(0.10968169942141635),
          np.float64(0.10968169942141635),
          np.float64(0.10882143751650175),
          np.float64(0.10882143751650175),
          np.float64(0.10882143751650175),
          np.float64(0.10882143751650175),
          np.float64(0.10882143751650175),
          np.float64(0.10882143751650175),
          np.float64(0.10882143751650175),
          np.float64(0.10838874619051501),
          np.float64(0.10838874619051501),
          np.float64(0.10838874619051501),
          np.float64(0.10838874619051501),
          np.float64(0.10838874619051501),
          np.float64(0.10814761408717502),
          np.float64(0.10814761408717502),
          np.float64(0.10814761408717502),
          np.float64(0.10814761408717502),
          np.float64(0.10814761408717502),
          np.float64(0.10814761408717502),
          np.float64(0.10814761408717502),
          np.float64(0.10743376064838502),
          np.float64(0.10743376064838502),
          np.float64(0.106676149412533),
          np.float64(0.106676149412533),
          np.float64(0.106676149412533),
          np.float64(0.106676149412533),
          np.float64(0.10650358071057624),
          np.float64(0.10650358071057624),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10619884881071831),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10526315789473685),
          np.float64(0.10390486669322622),
          np.float64(0.10385162309931599),
          np.float64(0.1037571695799112),
          np.float64(0.1037571695799112),
          np.float64(0.1037571695799112),
          np.float64(0.1037571695799112),
          np.float64(0.1037571695799112),
          np.float64(0.1037571695799112),
          np.float64(0.10301070542879115),
          np.float64(0.10259783520851541),
          np.float64(0.10259783520851541),
          np.float64(0.10259783520851541),
          np.float64(0.10259783520851541),
          np.float64(0.10203255357771539),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.10147651372376076),
          np.float64(0.101338918387628),
          np.float64(0.101338918387628),
          np.float64(0.101338918387628),
          np.float64(0.101338918387628),
          np.float64(0.101338918387628),
          np.float64(0.101338918387628),
          np.float64(0.1003911722115382),
          np.float64(0.10012523486435178),
          np.float64(0.10012523486435178),
          np.float64(0.10012523486435178),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.09933992677987831),
          np.float64(0.0989541391990587),
          np.float64(0.0989541391990587),
          np.float64(0.0989541391990587),
          np.float64(0.0989541391990587),
          np.float64(0.0989541391990587),
          np.float64(0.0989541391990587),
          np.float64(0.0989541391990587),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.0978231976089037),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09733285267845754),
          np.float64(0.09694584179118516),
          np.float64(0.0967301666813349),
          np.float64(0.09567297464698798),
          np.float64(0.09558988911273408),
          np.float64(0.09544271444636668),
          np.float64(0.09544271444636668),
          np.float64(0.09544271444636668),
          np.float64(0.09544271444636668),
          np.float64(0.09544271444636668),
          np.float64(0.09544271444636668),
          np.float64(0.09544271444636668),
          np.float64(0.09464970485606021),
          np.float64(0.09464970485606021),
          np.float64(0.09464970485606021),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09365858115816941),
          np.float64(0.09304036594559835),
          np.float64(0.09269795493186431),
          np.float64(0.09269795493186431),
          np.float64(0.09269795493186431),
          np.float64(0.09197090092274487),
          np.float64(0.09197090092274487),
          np.float64(0.09197090092274487),
          np.float64(0.09197090092274487),
          np.float64(0.09197090092274487),
          np.float64(0.09197090092274487),
          np.float64(0.09197090092274487),
          np.float64(0.09183979479633063),
          np.float64(0.09176629354822471),
          np.float64(0.09176629354822471),
          np.float64(0.09086217008485092),
          np.float64(0.09086217008485092),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09037128496931669),
          np.float64(0.09012301173931252),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.08998425413316952),
          np.float64(0.0891313035643173),
          np.float64(0.08903057122447033),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08885233166386385),
          np.float64(0.08830215713766959),
          np.float64(0.08830215713766959),
          np.float64(0.08830215713766959),
          np.float64(0.08830215713766959),
          np.float64(0.08830215713766959),
          np.float64(0.08830215713766959),
          np.float64(0.08830215713766959),
          np.float64(0.08749572785196143),
          np.float64(0.08749572785196143),
          np.float64(0.08749572785196143),
          np.float64(0.08749572785196143),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08740748201220976),
          np.float64(0.08671099695241201),
          np.float64(0.08671099695241201),
          np.float64(0.08671099695241201),
          np.float64(0.08671099695241201),
          np.float64(0.08671099695241201),
          np.float64(0.08671099695241201),
          np.float64(0.08671099695241201),
          np.float64(0.08646430798325933),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08603090020146065),
          np.float64(0.08594700851870801),
          np.float64(0.08594700851870801),
          np.float64(0.08594700851870801),
          np.float64(0.08520286456846099),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08471737420873576),
          np.float64(0.08447772061910234),
          np.float64(0.08447772061910234),
          np.float64(0.08447772061910234),
          np.float64(0.08447772061910234),
          np.float64(0.08447772061910234),
          np.float64(0.0837707816583391),
          np.float64(0.0837707816583391),
          np.float64(0.0837707816583391),
          np.float64(0.0837707816583391),
          np.float64(0.0837707816583391),
          np.float64(0.0837707816583391),
          np.float64(0.0837707816583391),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.08346223261119858),
          np.float64(0.0830812984794528),
          np.float64(0.0830812984794528),
          np.float64(0.08240856434303293),
          np.float64(0.08240856434303293),
          np.float64(0.08240856434303293),
          np.float64(0.08240856434303293),
          np.float64(0.08240856434303293),
          np.float64(0.08240856434303291),
          np.float64(0.08226127456606226),
          np.float64(0.08226127456606226),
          np.float64(0.08226127456606226),
          np.float64(0.08226127456606226),
          np.float64(0.08226127456606226),
          np.float64(0.08175191193132876),
          np.float64(0.08175191193132876),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08111071056538127),
          np.float64(0.08048436365855338),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.08000711205939975),
          np.float64(0.07987230638308718),
          np.float64(0.0792740035289418),
          np.float64(0.07894736842105264),
          np.float64(0.07894736842105264),
          np.float64(0.07894736842105264),
          np.float64(0.07894736842105264),
          np.float64(0.07894736842105264),
          np.float64(0.07894736842105264),
          np.float64(0.07894736842105264),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07868894753646337),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07792865001991967),
          np.float64(0.07768997059248714),
          np.float64(0.0770085628656519),
          np.float64(0.07694837640638656),
          np.float64(0.07694837640638656),
          np.float64(0.07694837640638656),
          np.float64(0.07694837640638656),
          np.float64(0.07694837640638656),
          np.float64(0.07694837640638656),
          np.float64(0.07694837640638656),
          np.float64(0.07664241743562072),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.07647191129018727),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.076004188790721),
          np.float64(0.07594632491572426),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826384),
          np.float64(0.07509392614826382),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07443229275647868),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.07421560439929402),
          np.float64(0.0734718358370645),
          np.float64(0.07336739820667779),
          np.float64(0.07336739820667779),
          np.float64(0.07336739820667779),
          np.float64(0.07336739820667779),
          np.float64(0.07336739820667779),
          np.float64(0.07300534327409847),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100118),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07254762501100116),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07175473098524099),
          np.float64(0.07165743639000186),
          np.float64(0.07098727864204515),
          np.float64(0.07098727864204515),
          np.float64(0.07098727864204515),
          np.float64(0.07098727864204515),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07079923254047887),
          np.float64(0.07024393586862707),
          np.float64(0.07024393586862707),
          np.float64(0.07024393586862707),
          np.float64(0.07024393586862707),
          np.float64(0.07024393586862707),
          np.float64(0.0695678476053965),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06952346619889824),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06917144638660747),
          np.float64(0.06904602208326344),
          np.float64(0.06882472016116853),
          np.float64(0.06882472016116853),
          np.float64(0.06882472016116853),
          np.float64(0.06882472016116853),
          np.float64(0.06882472016116853),
          np.float64(0.06882472016116853),
          np.float64(0.06882472016116853),
          np.float64(0.0687817449960909),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06814662756363819),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06765100914917384),
          np.float64(0.06748819059987714),
          np.float64(0.06748819059987714),
          np.float64(0.06684847767323797),
          np.float64(0.06684847767323797),
          np.float64(0.06684847767323797),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.0662266178532522),
          np.float64(0.06622661785325219),
          np.float64(0.06622661785325219),
          np.float64(0.06622661785325219),
          np.float64(0.06622661785325219),
          np.float64(0.06622661785325219),
          np.float64(0.06622661785325219),
          np.float64(0.06562179588897107),
          np.float64(0.06562179588897107),
          np.float64(0.06562179588897107),
          np.float64(0.06562179588897107),
          np.float64(0.065033247714309),
          np.float64(0.065033247714309),
          np.float64(0.065033247714309),
          np.float64(0.065033247714309),
          np.float64(0.065033247714309),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06488856845230502),
          np.float64(0.06446025638903101),
          np.float64(0.06446025638903101),
          np.float64(0.06446025638903101),
          np.float64(0.06390214842634574),
          np.float64(0.06390214842634574),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06362847629757779),
          np.float64(0.06335829046432676),
          np.float64(0.06335829046432676),
          np.float64(0.06335829046432676),
          np.float64(0.06282808624375433),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.06243905410544627),
          np.float64(0.0623109738595896),
          np.float64(0.0623109738595896),
          np.float64(0.0623109738595896),
          np.float64(0.061806423257274694),
          np.float64(0.061806423257274694),
          np.float64(0.061806423257274694),
          np.float64(0.061806423257274694),
          np.float64(0.06158950357943867),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.06131393394849658),
          np.float64(0.060833032924035954),
          np.float64(0.060833032924035954),
          np.float64(0.060833032924035954),
          np.float64(0.060363272743915036),
          np.float64(0.060363272743915036),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.0602475233128778),
          np.float64(0.05990422978731538),
          np.float64(0.0597347691089499),
          np.float64(0.05945550264670635),
          np.float64(0.05945550264670635),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05923488777590923),
          np.float64(0.05901671065234752),
          np.float64(0.058989607683913654),
          np.float64(0.058587492514838024),
          np.float64(0.05850805181048254),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.0582716546748065),
          np.float64(0.05775642214923893),
          np.float64(0.05735393346764045),
          np.float64(0.05735393346764045),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.057353933467640436),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05647824947249051),
          np.float64(0.05619514869490164),
          ...]
```

In [71]:
```
def recommend(movie):
    movie_index = new_df[new_df['title'] == movie].index[0]
    distances = similarity[movie_index]
    return
```

In [72]:
```
new_df['title'] == 'Avatar'
```

Out[72]:
```
0        True
1       False
2       False
3       False
4       False
        ...
4804    False
4805    False
4806    False
4807    False
4808    False
Name: title, Length: 4806, dtype: bool
```

In [73]:
```
new_df[new_df['title'] == 'Avatar']
```

Out[73]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...

In [74]:
```
new_df[new_df['title'] == 'Batman Begins'].index[0]
```

Out[74]:
```
np.int64(119)
```

In [75]:
```
sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]
```

Out[75]:
```
[(539, np.float64(0.25038669783359574)),
 (1192, np.float64(0.24779731389167606)),
 (507, np.float64(0.24283093212859141)),
 (260, np.float64(0.2409900932515112)),
 (1214, np.float64(0.23939494881986934))]
```

In [76]:
```
def recommend(movie):
    movie_index = new_df[new_df['title'] == movie].index[0]
    distances = similarity[movie_index]
    movies_list = sorted(list(enumerate(distances)),reverse = True, key = lambda x:x[1])[1:6]

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

In [77]:
```
recommend('Batman Begins')
```

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

In [78]:
```
new_df.iloc[1216].title
```

Out[78]:
```
'Autumn in New York'
```

In [79]:
```
import pickle
```

In [80]:

```
pickle.dump(new_df, open('movies.pkl','wb'))
```

In [81]:

```
new_df['title'].values
```

Out[81]:

```
array(['Avatar', "Pirates of the Caribbean: At World's End", 'Spectre',
       ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
       'My Date with Drew'], dtype=object)
```

In [82]:

```
pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

In [83]:

```
pickle.dump(similarity,open('similarity.pkl','wb'))
```

In [ ]:

```

```