

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import osmovies.iloc[0]['genres']
```

```
In [2]: movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

```
In [3]: movies.head(2)
```

	budget	genres	homepage	id	keywords	original
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	["id": 1463, "name": "culture clash"], {"id": 1463, "name": "culture clash"}]	Avatar
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://disney.go.com/disneypictures/pirates/	285	["id": 270, "name": "ocean"], {"id": 726, "name": "pirates"}]	Pirates of the Caribbean: The Curse of the Black Pearl

```
In [4]: credits.head(2)
```

	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..."}]

```
In [5]: movies.shape
```

Out[5]: (4803, 20)

```
In [6]: credits.shape
```

```
Out[6]: (4803, 4)
```

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

```
In [9]: movies.head(2)
```

```
Out[9]:
```

	budget	genres	homepage	id	keywords	original
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "ocean"}]	

2 rows × 23 columns



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

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

```
In [11]: movies.columns
```

```
Out[11]: Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',  
                'original_title', 'overview', 'popularity', 'production_companies',  
                'production_countries', 'release_date', 'revenue', 'runtime',  
                'spoken_languages', 'status', 'tagline', 'title', 'vote_average',  
                'vote_count', 'movie_id', 'cast', 'crew'],  
              dtype='object')
```

```
In [12]: movies = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]
```

```
In [13]: movies.head()
```

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

```
In [14]: movies.shape
```

```
Out[14]: (4809, 7)
```

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

```
Out[15]: movie_id    0
title            0
overview         3
genres           0
keywords         0
cast             0
crew            0
dtype: int64
```

```
In [16]: movies.dropna(inplace=True)
```

```
In [17]: movies.shape
```

```
Out[17]: (4806, 7)
```

```
In [18]: movies.duplicated().sum()
```

```
Out[18]: 0
```

```
In [19]: movies.head(2)
```

```
Out[19]:
```

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	[{"id": 1463, "name": "culture clash"}, {"id": 1463, "name": "culture clash"}]	[{"cast_id": 242, "character": "Jake Sully", "name": "Jake Sully"}]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "na..."}]	[{"cast_id": 4, "character": "Captain Jack Spa...", "name": "Captain Jack Spa..."}]

```
In [20]: movies.iloc[0]['genres']
```

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

```
In [35]: import ast

def convert(text):
    l = []
    for i in ast.literal_eval(text):
        l.append(i['name'])

    return l
```

```
In [22]: movies['genres'] = movies['genres'].apply(convert)
```

```
In [23]: movies.head(2)
```

```
Out[23]:
```

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id": 1463, "name": "culture clash"}]	[{"cast_id": 242, "character": "Jake Sully", "name": "Jake Sully"}]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "na..."}]	[{"cast_id": 4, "character": "Captain Jack Spa...", "name": "Captain Jack Spa..."}]

```
In [24]: movies.iloc[0]['keywords']
```

```
Out[24]: '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "society"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"id": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "alien"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien planet"}, {"id": 10987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldier"}, {"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165431, "name": "anti war"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "name": "mind and soul"}, {"id": 209714, "name": "3d"}]'
```

```
In [25]: movies['keywords'] = movies['keywords'].apply(convert)
movies.head(2)
```

```
Out[25]:
```

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[{"cast_id": 242, "character": "Jake Sully", "..."}]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[{"cast_id": 4, "character": "Captain Jack Spa..."}]

```
In [26]: movies.iloc[0]['cast']
```

Out[26]: '[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington", "order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana", "order": 1}, {"cast_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52fe48009251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quaritch", "credit_id": "52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang", "order": 3}, {"cast_id": 5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "name": "Michelle Rodriguez", "order": 4}, {"cast_id": 8, "character": "Selfridge", "credit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovanni Ribisi", "order": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id": "52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "name": "Joel David Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_id": "52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder", "order": 7}, {"cast_id": 11, "character": "Eytukan", "credit_id": "52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "name": "Wes Studi", "order": 8}, {"cast_id": 10, "character": "Tsu\\'Tey", "credit_id": "52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso", "order": 9}, {"cast_id": 12, "character": "Dr. Max Patel", "credit_id": "52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "name": "Dileep Rao", "order": 10}, {"cast_id": 13, "character": "Lyle Wainfleet", "credit_id": "52fe48009251416c750ac9f5", "gender": 2, "id": 98215, "name": "Matt Gerald", "order": 11}, {"cast_id": 32, "character": "Private Fike", "credit_id": "52fe48009251416c750aca5b", "gender": 2, "id": 154153, "name": "Sean Anthony Moran", "order": 12}, {"cast_id": 33, "character": "Cryo Vault Med Tech", "credit_id": "52fe48009251416c750aca5f", "gender": 2, "id": 397312, "name": "Jason Whyte", "order": 13}, {"cast_id": 34, "character": "Venture Star Crew Chief", "credit_id": "52fe48009251416c750aca63", "gender": 2, "id": 42317, "name": "Scott Lawrence", "order": 14}, {"cast_id": 35, "character": "Lock Up Trooper", "credit_id": "52fe48009251416c750aca67", "gender": 2, "id": 986734, "name": "Kelly Kilgour", "order": 15}, {"cast_id": 36, "character": "Shuttle Pilot", "credit_id": "52fe48009251416c750aca6b", "gender": 0, "id": 1207227, "name": "James Patrick Pitt", "order": 16}, {"cast_id": 37, "character": "Shuttle Co-Pilot", "credit_id": "52fe48009251416c750aca6f", "gender": 0, "id": 1180936, "name": "Sean Patrick Murphy", "order": 17}, {"cast_id": 38, "character": "Shuttle Crew Chief", "credit_id": "52fe48009251416c750aca73", "gender": 2, "id": 1019578, "name": "Peter Dillon", "order": 18}, {"cast_id": 39, "character": "Tractor Operator / Troupe", "credit_id": "52fe48009251416c750aca77", "gender": 0, "id": 91443, "name": "Kevin Dorman", "order": 19}, {"cast_id": 40, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7b", "gender": 2, "id": 173391, "name": "Kelson Henderson", "order": 20}, {"cast_id": 41, "character": "Dragon Gunship Gunner", "credit_id": "52fe48009251416c750aca7f", "gender": 0, "id": 1207236, "name": "David Van Horn", "order": 21}, {"cast_id": 42, "character": "Dragon Gunship Navigator", "credit_id": "52fe48009251416c750aca83", "gender": 0, "id": 215913, "name": "Jacob Tomuri", "order": 22}, {"cast_id": 43, "character": "Suit #1", "credit_id": "52fe48009251416c750aca87", "gender": 0, "id": 143206, "name": "Michael Blain-Rozgay", "order": 23}, {"cast_id": 44, "character": "Suit #2", "credit_id": "52fe48009251416c750aca8b", "gender": 2, "id": 169676, "name": "Jon Curry", "order": 24}, {"cast_id": 46, "character": "Ambient Room Tech", "credit_id": "52fe48009251416c750aca8f", "gender": 0, "id": 1048610, "name": "Luke Hawker", "order": 25}, {"cast_id": 47, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48009251416c750aca93", "gender": 0, "id": 42288, "name": "Woodly Schultz", "order": 26}, {"cast_id": 48, "character": "Horse Clan Leader", "credit_id": "52fe48009251416c750aca97", "gender": 2, "id": 68278, "name": "Peter Mensah", "order": 27}, {"cast_id": 49, "character": "Link Room Tech", "credit_id": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247, "name": "Sonia Yee", "order": 28}, {"cast_id": 50, "character": "Basket

```

ball Avatar / Troupe", "credit_id": "52fe48009251416c750aca9f", "gender":
1, "id": 1207248, "name": "Jahnel Curfman", "order": 29}, {"cast_id": 51,
"character": "Basketball Avatar", "credit_id": "52fe48009251416c750acaa3",
"gender": 0, "id": 89714, "name": "Ilram Choi", "order": 30}, {"cast_id":
52, "character": "Na\`vi Child", "credit_id": "52fe48009251416c750acaa7",
"gender": 0, "id": 1207249, "name": "Kyla Warren", "order": 31}, {"cast_i
d": 53, "character": "Troupe", "credit_id": "52fe48009251416c750acaab", "g
ender": 0, "id": 1207250, "name": "Lisa Roumain", "order": 32}, {"cast_i
d": 54, "character": "Troupe", "credit_id": "52fe48009251416c750acaaf", "g
ender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id":
57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gende
r": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 55,
"character": "Troupe", "credit_id": "52fe48009251416c750acab3", "gender":
0, "id": 1207251, "name": "Taylor Kibby", "order": 35}, {"cast_id": 56, "c
haracter": "Troupe", "credit_id": "52fe48009251416c750acab7", "gender": 0,
"id": 1207252, "name": "Jodie Landau", "order": 36}, {"cast_id": 58, "char
acter": "Troupe", "credit_id": "52fe48009251416c750acabf", "gender": 0, "i
d": 1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "characte
r": "Troupe", "credit_id": "52fe48009251416c750acac3", "gender": 0, "id":
1207257, "name": "Cullen B. Madden", "order": 38}, {"cast_id": 60, "charac
ter": "Troupe", "credit_id": "52fe48009251416c750acac7", "gender": 0, "i
d": 1207259, "name": "Joseph Brady Madden", "order": 39}, {"cast_id": 61,
"character": "Troupe", "credit_id": "52fe48009251416c750acacb", "gender":
0, "id": 1207262, "name": "Frankie Torres", "order": 40}, {"cast_id": 62,
"character": "Troupe", "credit_id": "52fe48009251416c750acacf", "gender":
1, "id": 1158600, "name": "Austin Wilson", "order": 41}, {"cast_id": 63,
"character": "Troupe", "credit_id": "52fe48019251416c750acad3", "gender":
1, "id": 983705, "name": "Sara Wilson", "order": 42}, {"cast_id": 64, "cha
racter": "Troupe", "credit_id": "52fe48019251416c750acad7", "gender": 0,
"id": 1207263, "name": "Tamica Washington-Miller", "order": 43}, {"cast_i
d": 65, "character": "Op Center Staff", "credit_id": "52fe48019251416c750a
cadb", "gender": 1, "id": 1145098, "name": "Lucy Briant", "order": 44},
{"cast_id": 66, "character": "Op Center Staff", "credit_id": "52fe48019251
416c750acadf", "gender": 2, "id": 33305, "name": "Nathan Meister", "orde
r": 45}, {"cast_id": 67, "character": "Op Center Staff", "credit_id": "52f
e48019251416c750acae3", "gender": 0, "id": 1207264, "name": "Gerry Blair",
"order": 46}, {"cast_id": 68, "character": "Op Center Staff", "credit_id":
"52fe48019251416c750acae7", "gender": 2, "id": 33311, "name": "Matthew Cha
mberlain", "order": 47}, {"cast_id": 69, "character": "Op Center Staff",
"credit_id": "52fe48019251416c750acaeb", "gender": 0, "id": 1207265, "nam
e": "Paul Yates", "order": 48}, {"cast_id": 70, "character": "Op Center Du
ty Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 1
207266, "name": "Wray Wilson", "order": 49}, {"cast_id": 71, "character":
"Op Center Staff", "credit_id": "52fe48019251416c750acaf3", "gender": 2,
"id": 54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "charac
ter": "Dancer", "credit_id": "52fe48019251416c750acaf7", "gender": 0, "i
d": 1207267, "name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 7
3, "character": "Dancer", "credit_id": "52fe48019251416c750acafb", "gende
r": 0, "id": 1207268, "name": "Carvon Futrell", "order": 52}, {"cast_id":
74, "character": "Dancer", "credit_id": "52fe48019251416c750acaff", "gende
r": 0, "id": 1207269, "name": "Brandon Jelkes", "order": 53}, {"cast_id":
75, "character": "Dancer", "credit_id": "52fe48019251416c750acb03", "gende
r": 0, "id": 1207270, "name": "Micah Moch", "order": 54}, {"cast_id": 76,
"character": "Dancer", "credit_id": "52fe48019251416c750acb07", "gender":
0, "id": 1207271, "name": "Hanniyah Muhammad", "order": 55}, {"cast_id": 7
7, "character": "Dancer", "credit_id": "52fe48019251416c750acb0b", "gende
r": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_i
d": 78, "character": "Dancer", "credit_id": "52fe48019251416c750acb0f", "g
ender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_i
d": 79, "character": "Dancer", "credit_id": "52fe48019251416c750acb13", "g
ender": 0, "id": 1207274, "name": "April Marie Thomas", "order": 58}, {"ca

```


st_id": 80, "character": "Dancer", "credit_id": "52fe48019251416c750acb17", "gender": 0, "id": 1207275, "name": "Bravita A. Threatt", "order": 59}, {"cast_id": 81, "character": "Mining Chief (uncredited)", "credit_id": "52fe48019251416c750acb1b", "gender": 0, "id": 1207276, "name": "Colin Bleasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (uncredited)", "credit_id": "52fe48019251416c750acb1f", "gender": 0, "id": 107969, "name": "Mike Bodnar", "order": 61}, {"cast_id": 83, "character": "Richard (uncredited)", "credit_id": "52fe48019251416c750acb23", "gender": 0, "id": 1207278, "name": "Matt Clayton", "order": 62}, {"cast_id": 84, "character": "Nav'i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gender": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "order": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2, "id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 14, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 90, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 91, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 92, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 93, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 94, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 95, "character": "Ops Centreworkeer (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 96, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 97, "character": "Col. Quaritch's Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "T. J. Storm", "order": 77}, {"cast_id": 98, "character": "Female Marine (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name": "Jodie Taylor", "order": 78}, {"cast_id": 99, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 100, "character": "Geologist (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207292, "name": "Richard Whiteside", "order": 80}, {"cast_id": 101, "character": "Na'vi (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Nikie Zambo", "order": 81}, {"cast_id": 102, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}]'

```
In [29]: def convert_cast(text):
1 = []
counter = 0
for i in ast.literal_eval(text):
    if counter < 3:
        l.append(i['name'])
        counter +=1

return l
```

```
In [30]: movies['cast'] = movies['cast'].apply(convert_cast)
movies.head(2)
```

```
Out[30]:
```

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]

```
In [31]: movies.iloc[0]['crew']
```

```
{
  "credit_id": "5495a12f92514130fc002e94",
  "department": "Crew",
  "gender": 0,
  "id": 1357071,
  "job": "Transportation Coordinator",
  "name": "Denny Cairra",
  "credit_id": "5495976fc3a36811530026b0",
  "department": "Sound",
  "gender": 0,
  "id": 1360103,
  "job": "Supervising Sound Editor",
  "name": "Addison Teague",
  "credit_id": "54491837c3a3680fb1001c5a",
  "department": "Art",
  "gender": 2,
  "id": 1376887,
  "job": "Set Designer",
  "name": "C. Scott Baker",
  "credit_id": "54491878c3a3680fb4001c9d",
  "department": "Art",
  "gender": 0,
  "id": 1376888,
  "job": "Set Designer",
  "name": "Luke Caska",
  "credit_id": "544918dac3a3680fa5001ae0",
  "department": "Art",
  "gender": 0,
  "id": 1376889,
  "job": "Set Designer",
  "name": "David Chow",
  "credit_id": "544919110e0a267486001b68",
  "department": "Art",
  "gender": 0,
  "id": 1376890,
  "job": "Set Designer",
  "name": "Jonathan Dyer",
  "credit_id": "54491967c3a3680faa001b5e",
  "department": "Art",
  "gender": 0,
  "id": 1376891,
  "job": "Set Designer",
  "name": "Joseph Hiura",
  "credit_id": "54491997c3a3680fb1001c8a",
  "department": "Art",
  "gender": 0,
  "id": 1376892,
  "job": "Art Department Coordinator",
  "name": "Rebecca Jellie",
  "credit_id": "544919ba0e0a26748f001b42",
  "department": "Art",
  "gender": 0,
  "id": 1376893,
  "job": "Set Designer",
  "name": "Robert Andrew Johnson",
  "credit_id": "54491b1dc3a3680faa001b8c"
}
```

```
In [32]: def fetch_director(text):
          l = []
          for i in ast.literal_eval(text):
              if i['job'] == 'Director':
                  l.append(i['name'])
                  break

          return l
```

```
In [33]: movies['crew'] = movies['crew'].apply(fetch_director)
```

```
In [34]: movies.head(2)
```

```
Out[34]:
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]

```
In [36]: movies.iloc[0]['overview']
```

```
Out[36]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization.'
```

```
In [37]: movies['overview'] = movies['overview'].apply(lambda x:x.split())
movies.head()
```

```
Out[37]:
```

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

```
In [39]: #Sam Worthington
#SamWorthington

def remove_space(word):
    l = []
    for i in word:
        l.append(i.replace(" ", ""))
    return l
```

```
In [40]: movies['cast'] = movies['cast'].apply(remove_space)
movies['crew'] = movies['crew'].apply(remove_space)
movies['genres'] = movies['genres'].apply(remove_space)
movies['keywords'] = movies['keywords'].apply(remove_space)
```

```
In [41]: movies.head()
```

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

```
In [42]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords']
```

```
In [43]: movies.head(2)
```

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]

```
In [44]: movies.iloc[0]['tags']
```

```
Out[44]: ['In',  
          'the',  
          '22nd',  
          'century,',  
          'a',  
          'paraplegic',  
          'Marine',  
          'is',  
          'dispatched',  
          'to',  
          'the',  
          'moon',  
          'Pandora',  
          'on',  
          'a',  
          'unique',  
          'mission,',  
          'but',  
          'becomes',  
          'torn',  
          'between',  
          'following',  
          'orders',  
          'and',  
          'protecting',  
          'an',  
          'alien',  
          'civilization.',  
          'Action',  
          'Adventure',  
          'Fantasy',  
          'ScienceFiction',  
          'cultureclash',  
          'future',  
          'spacewar',  
          'spacecolony',  
          'society',  
          'spacetravel',  
          'futuristic',  
          'romance',  
          'space',  
          'alien',  
          'tribe',  
          'alienplanet',  
          'cgi',  
          'marine',  
          'soldier',  
          'battle',  
          'loveaffair',  
          'antiwar',  
          'powerrelations',  
          'mindandsoul',  
          '3d',  
          'SamWorthington',  
          'ZoeSaldana',  
          'SigourneyWeaver',  
          'JamesCameron']
```

```
In [45]: new_df = movies[['movie_id', 'title', 'tags']]
```

```
In [46]: new_df.head()
```

```
Out[46]:
```

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

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

C:\Users\AKHIL\AppData\Local\Temp\ipykernel_13244\1955922539.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

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

```
Out[47]:
```

	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 [48]: new_df.iloc[0]['tags']
```

```
Out[48]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization. Action Adventure Fantasy ScienceFiction culture clash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCameron'
```

```
In [49]: new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
new_df.head()
```

C:\Users\AKHIL\AppData\Local\Temp\ipykernel_13244\842480936.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

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

```
Out[49]:
```

	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 [50]: new_df.iloc[0]['tags']
```

```
Out[50]: 'in the 22nd century, a paraplegic marine is dispatched to the moon pandor
a on a unique mission, but becomes torn between following orders and prote
cting an alien civilization. action adventure fantasy sciencefiction cultu
reclash future spacewar spacecolony society spacetravel futuristic romance
space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar
powerrelations mindandsoul 3d samworthington zoesaldana sigourneyweaver ja
mescameron'
```

```
In [51]: import nltk
from nltk.stem import PorterStemmer
```

```
In [52]: ps = PorterStemmer()
```

```
In [53]: def stems(text):
    l = []
    for i in text.split():
        l.append(ps.stem(i))

    return " ".join(l)
```



```
In [54]: new_df['tags'] = new_df['tags'].apply(stems)
```

C:\Users\AKHIL\AppData\Local\Temp\ipykernel_13244\3973021881.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

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

```
In [55]: new_df.iloc[0]['tags']
```

```
Out[55]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on  
a uniqu mission, but becom torn between follow order and protect an alien  
civilization. action adventur fantasi sciencefict cultureclash futur space  
war spacecoloni societi spacetravel futurist romanc space alien tribe alie  
nplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3d  
samworthington zoesaldana sigourneyweav jamescameron'
```

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

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

```
In [58]: vector
```

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

```
In [59]: vector.shape
```

```
Out[59]: (4806, 5000)
```

```
In [60]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [61]: similary = cosine_similarity(vector)
```

```
In [62]: similary
```

```
Out[62]: array([[1.          , 0.08346223, 0.0860309 , ..., 0.04499213, 0.          ,
                0.          ],
               [0.08346223, 1.          , 0.06063391, ..., 0.02378257, 0.          ,
                0.02615329],
               [0.0860309 , 0.06063391, 1.          , ..., 0.02451452, 0.          ,
                0.          ],
               ...,
               [0.04499213, 0.02378257, 0.02451452, ..., 1.          , 0.03962144,
                0.04229549],
               [0.          , 0.          , 0.          , ..., 0.03962144, 1.          ,
                0.08714204],
               [0.          , 0.02615329, 0.          , ..., 0.04229549, 0.08714204,
                1.          ]])
```

```
In [63]: similary.shape
```

```
Out[63]: (4806, 4806)
```

```
In [64]: new_df[new_df['title'] == 'Spider-Man'].index[0]
```

```
Out[64]: 159
```

```
In [67]: def recommend(movie):
          index = new_df[new_df['title'] == movie].index[0]
          distances = sorted(list(enumerate(similary[index])), reverse=True , key
          for i in distances[1:6]:
              print(new_df.iloc[i[0]].title)
```

```
In [68]: recommend('The Dark Knight Rises')
```

```
The Dark Knight
Batman Returns
Batman
Batman Forever
Batman Begins
```

```
In [70]: import pickle
```

```
pickle.dump(new_df,open('artificats/movie_list.pkl', 'wb'))
pickle.dump(similary,open('artificats/similarity.pkl', 'wb'))
```

```
In [71]: new_df['title'].values
```

```
Out[71]: array(['Avatar', 'Pirates of the Caribbean: At World's End', 'Spectre',
                ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
                'My Date with Drew'], dtype=object)
```

```
In [73]: pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

```
In [ ]:
```

