Content Based movie recommendation system

In [25]: ▶|
```python
import numpy as np
import pandas as pd
```

In [82]: ▶|
```python
movies = pd.read_csv("Datasets/tmdb_5000_movies.csv")
credits = pd.read_csv("Datasets/tmdb_5000_credits.csv")
movies.shape
credits.shape
```

Out[82]: (4803, 4)

In [67]: ▶|
```python
movies.head()
```

Out[67]:

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

```
In [83]:  ▶| credits.head()
```

Out[83]:

	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 [84]:  ▶| #movies.isnull().sum()
```

```
In [85]:  ▶| credits.head(1)
```

Out[85]:

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

```
In [86]:  ▶| credits.head(1)["cast"]
```

Out[86]: 0   [{"cast_id": 242, "character": "Jake Sully", "...
          Name: cast, dtype: object

```
In [87]:  ▶| credits.head()["cast"].values
```

Out[87]: array(['[{"cast_id": 242, "character": "Jake Sully", "credit_id": "560
2a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthin
gton", "order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_i
d": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe
Saldana", "order": 1}, {"cast_id": 25, "character": "Dr. Grace Augusti
ne", "credit_id": "52fe48009251416c750aca39", "gender": 1, "id": 1020
5, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "characte
r": "Col. Quaritch", "credit_id": "52fe48009251416c750ac9cf", "gende
r": 2, "id": 32747, "name": "Stephen Lang", "order": 3}, {"cast_id":
5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d
3", "gender": 1, "id": 17647, "name": "Michelle Rodriguez", "order":
4}, {"cast_id": 8, "character": "Selfridge", "credit_id": "52fe4800925
1416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovanni Ribisi", "o
rder": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id":
"52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "name": "Joel Da
vid Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_i
d": "52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH
Pounder", "order": 7}, {"cast_id": 11, "character": "Eytukan", "credit
_id": "52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "name": "W
```

## Merging both datasets

In [ ]:  ▶|

In [50]:  ▶| movies.shape

Out[50]: (4806, 7)

In [51]:  ▶| credits.shape

Out[51]: (4803, 4)

In [88]:  ▶| movies = movies.merge(credits, on = "title")

In [89]:  ▶| movies.shape

Out[89]: (4809, 23)
```

```
In [90]:  ▶| movies.head(5)
```

Out[90]:

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

5 rows × 23 columns

◄ ▬▬▬▬▬▬▬▬▬▬ ►

```
In [91]:  ▶| # genres
             # id
             # keywords
             # title
             # overview
             # cast
             # crew
```

```
In [93]:  ▶| movies = movies[["movie_id", "title", "overview" , "genres", "keywords", "
```

`movies.head()`

Out[94]:

	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", "...	[{" "52fe48009251416c7...
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...	[{" "52fe4232c3a36847f...
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...	[{" "54805967c3a36829b...
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...	[{" "52fe4781c3a36847f...
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...	[{" "52fe479ac3a36847f...

Data Preprocessing

In [95]: `movies.isnull().sum()`

Out[95]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

```
In [96]:  ▶|  movies.dropna(inplace = True)
```

```
In [97]:  ▶|  movies.isnull().sum()
```

```
Out[97]:  movie_id    0
          title       0
          overview    0
          genres      0
          keywords    0
          cast        0
          crew        0
          dtype: int64
```

```
In [98]:  ▶|  movies.duplicated().sum()
```

```
Out[98]:  0
```

```
In [99]:  ▶|  movies.iloc[0]
```

```
Out[99]:  movie_id                                         19995
          title                                           Avatar
          overview    In the 22nd century, a paraplegic Marine is di...
          genres      [{"id": 28, "name": "Action"}, {"id": 12, "nam...
          keywords    [{"id": 1463, "name": "culture clash"}, {"id":...
          cast        [{"cast_id": 242, "character": "Jake Sully", "...
          crew        [{"credit_id": "52fe48009251416c750aca23", "de...
          Name: 0, dtype: object
```

```
In [100]:  ▶|  movies.iloc[0].genres
```

```
Out[100]:  '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 1
           4, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [21]:  ▶|  # converting the genres in list
```

```
In [105]:  ▶|  import ast
              def convert(obj):
                  l = []
                  for i in ast.literal_eval(obj):
                      l.append(i["name"])

                  return l
```

```
In [106]:  ▶|  convert('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {
```

```
Out[106]:  ['Action', 'Adventure', 'Fantasy', 'Science Fiction']
```

```
In [107]:  ▶| movies["genres"] = movies["genres"].apply(convert)
```

```
In [108]:  ▶| movies.head()
```

Out[108]:

	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":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"cr "52fe48009251416c75(
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"cr "52fe4232c3a36847f8(
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"cr "54805967c3a36829b5(
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"cr "52fe4781c3a36847f81
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"cr "52fe479ac3a36847f81

```
In [109]:  ▶| movies["keywords"] = movies["keywords"].apply(convert)
```

	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", "...	[{"cr "52fe48009251416c75(
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...	[{"cr "52fe4232c3a36847f8(
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"cr "54805967c3a36829b5(
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"cr "52fe4781c3a36847f81
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[{"cast_id": 5, "character": "John Carter", "c...	[{"cr "52fe479ac3a36847f81

```
In [111]: ▶| movies["cast"][0]
```

Out[111]: '[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a36
85532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington", "orde
r": 0}, {"cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251
416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana", "order":
1}, {"cast_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52f
e48009251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney Weav
er", "order": 2}, {"cast_id": 4, "character": "Col. Quaritch", "credit_i
d": "52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephe
n Lang", "order": 3}, {"cast_id": 5, "character": "Trudy Chacon", "credit
_id": "52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "name": "Mich
elle Rodriguez", "order": 4}, {"cast_id": 8, "character": "Selfridge", "c
redit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name":
"Giovanni Ribisi", "order": 5}, {"cast_id": 7, "character": "Norm Spellma
n", "credit_id": "52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "n
ame": "Joel David Moore", "order": 6}, {"cast_id": 9, "character": "Moa
t", "credit_id": "52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "n
ame": "CCH Pounder", "order": 7}, {"cast_id": 11, "character": "Eytukan",
"credit_id": "52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "nam
e": "Wes Studi", "order": 8}, {"cast_id": 10, "character": "Tsu\'Tey", "c
redit_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, "nam
e": "Dileep Rao", "order": 10}, {"cast_id": 13, "character": "Lyle Wainfl
eet", "credit_id": "52fe48009251416c750ac9f5", "gender": 2, "id": 98215,
"name": "Matt Gerald", "order": 11}, {"cast_id": 32, "character": "Privat
e Fike", "credit_id": "52fe48009251416c750aca5b", "gender": 2, "id": 1541
53, "name": "Sean Anthony Moran", "order": 12}, {"cast_id": 33, "characte
r": "Cryo Vault Med Tech", "credit_id": "52fe48009251416c750aca5f", "gend
er": 2, "id": 397312, "name": "Jason Whyte", "order": 13}, {"cast_id": 3
4, "character": "Venture Star Crew Chief", "credit_id": "52fe48009251416c
750aca63", "gender": 2, "id": 42317, "name": "Scott Lawrence", "order": 1
4}, {"cast_id": 35, "character": "Lock Up Trooper", "credit_id": "52fe480
09251416c750aca67", "gender": 2, "id": 986734, "name": "Kelly Kilgour",
"order": 15}, {"cast_id": 36, "character": "Shuttle Pilot", "credit_id":
"52fe48009251416c750aca6b", "gender": 0, "id": 1207227, "name": "James Pa
trick 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": "Peter Dillon", "order": 18}, {"cast_id": 39,
"character": "Tractor Operator / Troupe", "credit_id": "52fe48009251416c7
50aca77", "gender": 0, "id": 91443, "name": "Kevin Dorman", "order": 19},
{"cast_id": 40, "character": "Dragon Gunship Pilot", "credit_id": "52fe48
009251416c750aca7b", "gender": 2, "id": 173391, "name": "Kelson Henderso
n", "order": 20}, {"cast_id": 41, "character": "Dragon Gunship Gunner",
"credit_id": "52fe48009251416c750aca7f", "gender": 0, "id": 1207236, "nam
e": "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, "char
acter": "Suit #1", "credit_id": "52fe48009251416c750aca87", "gender": 0,
"id": 143206, "name": "Michael Blain-Rozgay", "order": 23}, {"cast_id": 4
4, "character": "Suit #2", "credit_id": "52fe48009251416c750aca8b", "gend
er": 2, "id": 169676, "name": "Jon Curry", "order": 24}, {"cast_id": 46,
"character": "Ambient Room Tech", "credit_id": "52fe48009251416c750aca8
f", "gender": 0, "id": 1048610, "name": "Luke Hawker", "order": 25}, {"ca
st_id": 47, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe
48009251416c750aca93", "gender": 0, "id": 42288, "name": "Woody Schultz",

"order": 26}, {"cast_id": 48, "character": "Horse Clan Leader", "credit_id": "52fe48009251416c750aca97", "gender": 2, "id": 68278, "name": "Peter Mensah", "order": 27}, {"cast_id": 49, "character": "Link Room Tech", "credit_id": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247, "name": "Sonia Yee", "order": 28}, {"cast_id": 50, "character": "Basketball Avatar / Troupe", "credit_id": "52fe48009251416c750aca9f", "gender": 1, "id": 1207248, "name": "Jahnel Curfman", "order": 29}, {"cast_id": 51, "character": "Basketball Avatar", "credit_id": "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_id": 53, "character": "Troupe", "credit_id": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Roumain", "order": 32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52fe48009251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id": 57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 55, "character": "Troupe", "credit_id": "52fe48009251416c750acab3", "gender": 0, "id": 1207251, "name": "Taylor Kibby", "order": 35}, {"cast_id": 56, "character": "Troupe", "credit_id": "52fe48009251416c750acab7", "gender": 0, "id": 1207252, "name": "Jodie Landau", "order": 36}, {"cast_id": 58, "character": "Troupe", "credit_id": "52fe48009251416c750acabf", "gender": 0, "id": 1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "character": "Troupe", "credit_id": "52fe48009251416c750acac3", "gender": 0, "id": 1207257, "name": "Cullen B. Madden", "order": 38}, {"cast_id": 60, "character": "Troupe", "credit_id": "52fe48009251416c750acac7", "gender": 0, "id": 1207259, "name": "Joseph Brady Madden", "order": 39}, {"cast_id": 61, "character": "Troupe", "credit_id": "52fe48009251416c750acacb", "gender": 0, "id": 1207262, "name": "Frankie Torres", "order": 40}, {"cast_id": 62, "character": "Troupe", "credit_id": "52fe48009251416c750acacf", "gender": 1, "id": 1158600, "name": "Austin Wilson", "order": 41}, {"cast_id": 63, "character": "Troupe", "credit_id": "52fe48019251416c750acad3", "gender": 1, "id": 983705, "name": "Sara Wilson", "order": 42}, {"cast_id": 64, "character": "Troupe", "credit_id": "52fe48019251416c750acad7", "gender": 0, "id": 1207263, "name": "Tamica Washington-Miller", "order": 43}, {"cast_id": 65, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acadb", "gender": 1, "id": 1145098, "name": "Lucy Briant", "order": 44}, {"cast_id": 66, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acadf", "gender": 2, "id": 33305, "name": "Nathan Meister", "order": 45}, {"cast_id": 67, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae3", "gender": 0, "id": 1207264, "name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae7", "gender": 2, "id": 33311, "name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaeb", "gender": 0, "id": 1207265, "name": "Paul Yates", "order": 48}, {"cast_id": 70, "character": "Op Center Duty Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 1207266, "name": "Wray Wilson", "order": 49}, {"cast_id": 71, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaf3", "gender": 2, "id": 54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "character": "Dancer", "credit_id": "52fe48019251416c750acaf7", "gender": 0, "id": 1207267, "name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 73, "character": "Dancer", "credit_id": "52fe48019251416c750acafb", "gender": 0, "id": 1207268, "name": "Carvon Futrell", "order": 52}, {"cast_id": 74, "character": "Dancer", "credit_id": "52fe48019251416c750acaff", "gender": 0, "id": 1207269, "name": "Brandon Jelkes", "order": 53}, {"cast_id": 75, "character": "Dancer", "credit_i

d": "52fe48019251416c750acb03", "gender": 0, "id": 1207270, "name": "Micah Moch", "order": 54}, {"cast_id": 76, "character": "Dancer", "credit_id": "52fe48019251416c750acb07", "gender": 0, "id": 1207271, "name": "Hanniyah Muhammad", "order": 55}, {"cast_id": 77, "character": "Dancer", "credit_id": "52fe48019251416c750acb0b", "gender": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "credit_id": "52fe48019251416c750acb0f", "gender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_id": 79, "character": "Dancer", "credit_id": "52fe48019251416c750acb13", "gender": 0, "id": 1207274, "name": "April Marie Thomas", "order": 58}, {"cast_id": 80, "character": "Dancer", "credit_id": "52fe48019251416c750acb17", "gender": 0, "id": 1207275, "name": "Bravita A. Threatt", "order": 59}, {"cast_id": 81, "character": "Mining Chief (uncredited)", "credit_id": "52fe48019251416c750acb1b", "gender": 0, "id": 1207276, "name": "Colin Bleasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (uncredited)", "credit_id": "52fe48019251416c750acb1f", "gender": 0, "id": 107969, "name": "Mike Bodnar", "order": 61}, {"cast_id": 83, "character": "Richard (uncredited)", "credit_id": "52fe48019251416c750acb23", "gender": 0, "id": 1207278, "name": "Matt Clayton", "order": 62}, {"cast_id": 84, "character": "Nav\'i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gender": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "order": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2, "id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 14, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 90, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 91, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 92, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 93, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 94, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 95, "character": "Ops Centreworker (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 96, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 97, "character": "Col. Quaritch\'s Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "T. J. Storm", "order": 77}, {"cast_id": 98, "character": "Female Marine (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name": "Jodie Taylor", "order": 78}, {"cast_id": 99, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_

id": 100, "character": "Geologist (uncredited)", "credit_id": "52fe480192
51416c750acb67", "gender": 0, "id": 1207292, "name": "Richard Whiteside",
"order": 80}, {"cast_id": 101, "character": "Na\'vi (uncredited)", "credi
t_id": "52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Ni
kie Zambo", "order": 81}, {"cast_id": 102, "character": "Ambient Room Tec
h / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id":
42286, "name": "Julene Renee", "order": 82}]'

In [112]: ⏭
```python
# getting first three cast dictionaries
def convert1(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 [113]: ⏭
```python
movies["cast"] = movies["cast"].apply(convert1)
```

```
In [114]: ▶| movies.head()
```

Out[114]:

	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]	[{"c "52fe48009251416c75
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]	[{"c "52fe4232c3a36847f8
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[{"c "54805967c3a36829b5
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[{"c "52fe4781c3a36847f8
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[{"c "52fe479ac3a36847f8

```
In [115]:  ▶  movies["crew"][0]
```

```
Out[115]:  '[{"credit_id": "52fe48009251416c750aca23", "department": "Editing",
           "gender": 0, "id": 1721, "job": "Editor", "name": "Stephen E. Rivki
           n"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "g
           ender": 2, "id": 496, "job": "Production Design", "name": "Rick Carte
           r"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound",
           "gender": 0, "id": 900, "job": "Sound Designer", "name": "Christopher
           Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Soun
           d", "gender": 0, "id": 900, "job": "Supervising Sound Editor", "name":
           "Christopher Boyes"}, {"credit_id": "539c4a4cc3a36810c9002101", "depar
           tment": "Production", "gender": 1, "id": 1262, "job": "Casting", "nam
           e": "Mali Finn"}, {"credit_id": "5544ee3b925141499f0008fc", "departmen
           t": "Sound", "gender": 2, "id": 1729, "job": "Original Music Compose
           r", "name": "James Horner"}, {"credit_id": "52fe48009251416c750ac9c3",
           "department": "Directing", "gender": 2, "id": 2710, "job": "Director",
           "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac9d9", "d
           epartment": "Writing", "gender": 2, "id": 2710, "job": "Writer", "nam
           e": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "depar
           tment": "Editing", "gender": 2, "id": 2710, "job": "Editor", "name":
           "James Cameron"}, {"credit_id": "52fe48009251416c750aca29", "departmen
```

```
In [116]:  ▶  def get_director(obj):
                  l = []
                  for i in ast.literal_eval(obj):
                      if i["job"] == "Director":

                          l.append(i["name"])
                          break

                  return l
```

```
In [117]:  ▶  get_director('[{"credit_id": "52fe48009251416c750aca23", "department": "Ed
                ◀  ▬▬                                                                  ▶
```

```
Out[117]:  ['James Cameron']
```

```
In [118]:  ▶  movies["crew"] = movies["crew"].apply(get_director)
```

```
In [119]: ▶| movies.head()
```

Out[119]:

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

```
In [120]: ▶| movies["overview"][0]
```

Out[120]: '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 [121]: ▶| movies["overview"] = movies["overview"].apply(lambda x:x.split())
```

```
In [122]:  ▶| movies.head()
```

Out[122]:

	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 [123]:  ▶| # concatenating overview and genres keywords cast crew
```

```
In [124]:  ▶| movies["genres"] = movies["genres"].apply(lambda x: [i.replace(" ", "") fo
               movies["crew"] = movies["crew"].apply(lambda x:[i.replace(" ", "") for i i
               movies["keywords"] = movies["keywords"].apply(lambda x:[i.replace(" ", "")
               movies["cast"] = movies["cast"].apply(lambda x: [i.replace(" ", "") for i
```

```
In [125]:  ▶|  movies.head()
```

Out[125]:

	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]	[Jame
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]	[Gor
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[Sa
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[Christo
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[Andr

◀ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

```
In [126]:  ▶|  movies["tags"] = movies["overview"] + movies['genres'] + movies['keywords'
```

```
In [127]:  ▶|  movies.head()
```

Out[127]:

	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]	[Jame
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]	[Gor
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[Sa
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[Christo
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[Andr

```
In [128]:  ▶|  new_df = movies[["movie_id", "title", "tags"]]
```

```
In [129]:  ▶|  new_df
```

Out[129]:

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

4806 rows × 3 columns

```
In [130]:  ▶|  new_df["tags"] = new_df["tags"].apply(lambda x:" ".join(x))
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_12340\847019788.py:1: Setting
WithCopyWarning:
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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
  new_df["tags"] = new_df["tags"].apply(lambda x:" ".join(x))
```

```
In [131]:  ▶|  new_df
```

Out[131]:

	movie_id	title	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	A cryptic message from Bond's past sends him o...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...
4	49529	John Carter	John Carter is a war-weary, former military ca...
...	...	...	...
4804	9367	El Mariachi	El Mariachi just wants to play his guitar and ...
4805	72766	Newlyweds	A newlywed couple's honeymoon is upended by th...
4806	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...
4807	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...
4808	25975	My Date with Drew	Ever since the second grade when he first saw ...

4806 rows × 3 columns

```
In [132]:  ▶|  import nltk
               from nltk.stem.porter import PorterStemmer
               ps = PorterStemmer()
```

```
In [133]:  ▶|  def stem(text):
                   y = []
                   for i in text.split():
                       y.append(ps.stem(i))

                   return " ".join(y)
```

```
In [134]:  ▶| new_df["tags"] = new_df["tags"].apply(stem)
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_12340\825579503.py:1: Setting
WithCopyWarning:
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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
  new_df["tags"] = new_df["tags"].apply(stem)

```
In [135]:  ▶| new_df["tags"][0]
```

Out[135]: '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 spac
ewar spacecoloni societi spacetravel futurist romanc space alien tribe al
ienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul
3d samworthington zoesaldana sigourneyweav jamescameron'

```
In [ ]:  ▶|
```

```
In [136]:  ▶| new_df["tags"] = new_df["tags"].apply(lambda x:x.lower())
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_12340\3481706753.py:1: Settin
gWithCopyWarning:
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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
  new_df["tags"] = new_df["tags"].apply(lambda x:x.lower())

```
In [137]:  ▶| new_df
```

Out[137]:

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

4806 rows × 3 columns

```
In [ ]:  ▶|
```

Vectorization

converting texts into vectors
using bag of words technique

```
In [138]:  ▶| from sklearn.feature_extraction.text import CountVectorizer
              cv = CountVectorizer(max_features = 5000, stop_words= "english" )
```

```
In [139]:  ▶| vectors = cv.fit_transform(new_df["tags"]).toarray()
```

```
In [140]:  ▶| vectors[0]
```

Out[140]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)

```
In [141]:   ▶| cv.get_feature_names_out()

Out[141]:  array(['000', '007', '10', ..., 'zone', 'zoo', 'zooeydeschanel'],
               dtype=object)


In [ ]:    ▶|


In [142]:   ▶| from sklearn.metrics.pairwise import cosine_similarity
              similarity = cosine_similarity(vectors)


In [143]:   ▶| similarity.shape

Out[143]:  (4806, 4806)


In [144]:   ▶| sorted(list(enumerate(similarity[1])),reverse = True, key = lambda x:x[1])

Out[144]:  [(12, 0.4140043409440133),
            (199, 0.27500954910846337),
            (17, 0.2533201985524494),
            (216, 0.20579830217101058),
            (3572, 0.20579830217101058)]


In [145]:   ▶| 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 =

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


In [146]:   ▶| recommend("Batman")

              1363
              Batman
              210
              Batman & Robin
              119
              Batman Begins
              428
              Batman Returns
              4529
              The R.M.


In [ ]:    ▶|
```

In [ ]: