

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

```
movies = pd.read_csv('/content/tmdb_5000_movies.csv.zip')
credits = pd.read_csv('/content/tmdb_5000_credits.csv.zip')
```

```
movies.head()
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popu
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
1	300000000	[{"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
2	245000000	[{"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
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknighttrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112
4	260000000	[{"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

```
movies.shape
```

```
(4803, 20)
```

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

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

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

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

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	productio
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	[{"nam

1 rows × 23 columns

```
movies.head()
# budget
# homepage
# id
```

```
# original_language
# original_title
# popularity
# production_comapny
# production_countries
# release-date(not sure)
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popi
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"}]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150
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"}]	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name": "spy"}]	en	Spectre	A cryptic message from Bond's past sends him o...	107
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Fantasy"}]	http://www.thedarkknighttrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "dc comics"}]	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": 818, "name": "based on novel"}]	en	John Carter	John Carter is a war-weary, former military ca...	43
5 rows × 23 columns									

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

```
movies.head()
```

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

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

```

0
movie_id 0
title 0
overview 3
genres 0
keywords 0
cast 0
crew 0

```

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

```
import ast
```

```

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

```

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

```

movies['genres'] = movies['genres'].apply(convert)
movies.head()

```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id": ...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
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...	[{"credit_id": "54805967c3a36829b5002c41", "de...
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...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...

```

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

```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
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...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...

```

import ast
ast.literal_eval('{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Sc

```

```

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

```

```

def convert3(text):
    L = []
    counter = 0
    for i in ast.literal_eval(text):
        if counter < 3:
            L.append(i['name'])
        counter+=1
    return L

movies['cast'] = movies['cast'].apply(convert)
movies.head()

```

	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 Weave...	["credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley, ...	["credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...	["credit_id": "54805967c3a36829b5002c41", "de...
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, A...	["credit_id": "52fe4781c3a36847f81398c3", "de...

```
movies['cast'] = movies['cast'].apply(lambda x:x[0:3])
```

```

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

```

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

```
movies.sample(5)
```

	movie_id	title	overview	genres	keywords	cast	crew
279	280	Terminator 2: Judgment Day	Nearly 10 years have passed since Sarah Connor...	[Action, Thriller, Science Fiction]	[cyborg, shotgun, post-apocalyptic, dystopia, ...	[Arnold Schwarzenegger, Linda Hamilton, Robert...	[James Cameron]
2583	9424	Disturbing Behavior	Steve Clark (James Marsden) is a newcomer in t...	[Mystery, Horror, Science Fiction]	[suicide, sex, island, nightmare, chase, polic...	[James Marsden, Katie Holmes, Nick Stahl]	[David Nutter]
105	241259	Alice Through the Looking Glass	In the sequel to Tim Burton's "Alice in Wonder...	[Fantasy]	[based on novel, clock, queen, sequel, alice i...	[Johnny Depp, Mia Wasikowska, Anne Hathaway]	[James Bobin]
3417	46738	Incendies	A mother's last wishes send twins Jeanne and S...	[Drama, War, Mystery]	[prison, middle east, rape, muslim, militia, i...	[Lubna Azabal, Mélissa Désormeaux-Poulin, Maxi...	[Denis Villeneuve]

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

```
movies.head()
```

	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]

```

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])

```

```
movies.head()
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]
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]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]

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

```
new = movies.drop(columns=['overview', 'genres', 'keywords', 'cast', 'crew'])
```

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

	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...

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

```
vector = cv.fit_transform(new['tags']).toarray()
```

```
vector.shape
```

```
(4806, 5000)
```

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

```
similarity = cosine_similarity(vector)
```

```
similarity
```

```
array([[1.          , 0.08858079, 0.05905386, ..., 0.02478408, 0.02599376,
        0.          ],
       [0.08858079, 1.          , 0.06451613, ..., 0.02707652, 0.          ,
        0.          ],
       [0.05905386, 0.06451613, 1.          , ..., 0.02707652, 0.          ,
        0.          ],
       ...,
       [0.02478408, 0.02707652, 0.02707652, ..., 1.          , 0.07150969,
        0.0489116 ],
       [0.02599376, 0.          , 0.          , ..., 0.07150969, 1.          ,
        0.05129892],
       [0.          , 0.          , 0.          , ..., 0.0489116 , 0.05129892,
        1.          ]])
```

```
new[new['title'] == 'The Lego Movie'].index[0]
```

```
np.int64(744)
```

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

```
recommend('John Carter')
```

```
Star Trek: Insurrection
Mission to Mars
Captain America: The First Avenger
Escape from Planet Earth
Ghosts of Mars
```

```
import pickle
```

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

```
pickle.dump(new.to_dict(),open('movies_dict.pkl','wb'))
```

```
new['title'].values
```

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

```
new.to_dict()
```

Show hidden output