```
import numpy as np
import pandas as pd

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

movies = pd.read_csv('tmdb_5000_movies.csv')
credit = pd.read_csv('tmdb_5000_credits.csv')
```

movies.head(2)

3		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	30000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na	
4							•

movies.shape

→ (4803, 20)

credit.head()

0 19995 Avatar [{"cast_id": 242, "character": "Jake Sully", " [{"cast_id": 242, "52fe48009251416c750aca23", "de 1 285 Pirates of the Caribbean: At World's End [{"cast_id": 4, "character": "Captain Jack Spa "52fe4232c3a36847f800b579", "de 2 206647 Spectre "character": "James Bond", "cr [{"cast_id": 1, "54805967c3a36829b5002c41", "de 3 49026 The Dark Knight Rises [{"cast_id": 2, "character": "Bruce Wayne / Ba [{"credit_id": "52fe4781c3a36847f81398c3", "de	₹	r	movie_id	title	cast	crew
1 285 Caribbean: At World's End "character": "Captain Jack Spa "52fe4232c3a36847f800b579", "de "52fe4232c3a36847f800b579", "de "[{"credit_id": "52fe4232c3a36847f800b579", "de "52fe4232c3a36829b5002c41", "de "52f		0	19995	Avatar	"character": "Jake	
2 206647 Spectre "character": "James "54805967c3a36829b5002c41", Bond", "cr "de [{"cast_id": 2,		1	285	Caribbean: At	"character": "Captain	
3 49026 Rises "character": "Bruce "52fe4781c3a36847f81398c3" "de		2	206647	Spectre	"character": "James	"54805967c3a36829b5002c41",
		3	49026	•	"character": "Bruce	

movies.head()

_	budget	genres	homepage id	keywords	01
				[{"id"·	

	buuget	gein es	пошераве		KCywol us	0
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 r	ows × 23 colu	mns				
4						.

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

movies.head()

	cast	keywords	genres	overview	title	movie_id	
"52fe48009251416d	[{"cast_id": 242, "character": "Jake Sully", "	[{"id": 1463, "name": "culture clash"}, {"id":	[{"id": 28, "name": "Action"}, {"id": 12, "nam	In the 22nd century, a paraplegic Marine is di	Avatar	19995	0
"52fe4232c3a3684	[{"cast_id": 4, "character": "Captain Jack Spa	[{"id": 270, "name": "ocean"}, {"id": 726, "na	[{"id": 12, "name": "Adventure"}, {"id": 14, "	Captain Barbossa, long believed to be dead, ha	Pirates of the Caribbean: At World's End	285	1
"54805967c3a36829	[{"cast_id": 1, "character": "James Bond", "cr	[{"id": 470, "name": "spy"}, {"id": 818, "name	[{"id": 28, "name": "Action"}, {"id": 12, "nam	A cryptic message from Bond's past sends him o	Spectre	206647	2
+		[{"id"·					4

```
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()

7	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", "	[{' "52fe48009251416c7
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	[{" "52fe4232c3a36847f
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	[{" "54805967c3a36829b
					ll.,iЧ		
- 4							>

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

	cast	keywords	genres	overview	title	movie_id	
[{ "52fe48009251416c7	[{"cast_id": 242, "character": "Jake Sully", "	[culture clash, future, space war, space colon	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di	Avatar	19995	0
[{ "52fe4232c3a36847	[{"cast_id": 4, "character": "Captain Jack Spa	[ocean, drug abuse, exotic island, east india 	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha	Pirates of the Caribbean: At World's End	285	1
[{ "54805967c3a36829t	[{"cast_id": 1, "character": "James Bond", "cr	[spy, based on novel, secret agent, sequel,	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him	Spectre	206647	2
•		:					4

```
import ast
ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 87
```

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

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

7		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 Weave	["52fe48009251416c
	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,	["52fe4232c3a3684]
	2	206647	Spectre	A cryptic message from Bond's past sends him	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel,	[Daniel Craig, Christoph Waltz, Léa Seydoux, R	["54805967c3a3682\$
	4)

```
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['overview'] = movies['overview'].apply(lambda x:x.split())
movies.sample(5)

crew	cast	keywords	genres	overview	title	movie_id		₹
[Chris Wedge]	[Ray Romano, John Leguizamo, Denis Leary]	[human evolution, parents kids relationship, s	[Animation, Comedy, Family, Adventure]	[With, the, impending, ice, age, almost, upon,	Ice Age	425	631	
[Jerry Zucker]	[Sean Connery, Richard Gere, Julia Ormond]	[camelot, knight, king arthur, excalibur, knig	[Action, Adventure, Drama, Romance]	[The, timeless, tale, of, King, Arthur, and, t	First Knight	6520	535	
[Betty Thomas]	[Eddie Murphy, Owen Wilson, Famke	[budapest, kidnapping, boxer, secret	[Action, Adventure, Comedy, Thriller]	[When, the, Switchblade,, the, most,	I Spy	8427	596	

```
def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ",""))
movies['cast'] = movies['cast'].apply(collapse)
movies['crew'] = movies['crew'].apply(collapse)
movies['genres'] = movies['genres'].apply(collapse)
movies['keywords'] = movies['keywords'].apply(collapse)
movies.head()
₹
         movie_id
                         title overview
                                                                 keywords
                                                    genres
                                                                                         cast
                                    [In, the,
                                                              [cultureclash,
                                                    [Action,
                                     22nd,
                                                                             [SamWorthington,
                                                                    future.
                                                 Adventure,
      0
             19995
                         Avatar century,, a,
                                                                 spacewar,
                                                                                  ZoeSaldana,
                                                                                                 [Jar
                                                   Fantasy,
                                                               spacecolony,
                                 paraplegic,
                                                                            SigourneyWeaver]
                                            ScienceFiction]
                                    Marin...
                      Pirates of
                                  [Captain,
                                                                    [ocean,
                            the
                                 Barbossa,,
                                                [Adventure,
                                                                                [JohnnyDepp,
                                                                drugabuse,
      1
               285 Caribbean:
                                                   Fantasy,
                                                                                                  [G
                                                                                OrlandoBloom,
                                      long,
                                                               exoticisland,
                     At World's
                                  believed,
                                                    Action]
                                                                                KeiraKnightley]
                                                             eastindiatrad...
                                 to, be, d...
                           End
                                 [A, cryptic,
                                  message,
                                                                      [spy,
                                                                                 [DanielCraig,
                                                    [Action,
                                      from,
                                                             basedonnovel.
      2
            206647
                       Spectre
                                                 Adventure,
                                                                               ChristophWaltz,
                                    Bond's,
                                                               secretagent,
                                                                                 LéaSeydoux]
                                                    Crime]
                                      past,
                                                             sequel, mi6, ...
```

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

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

[dccomics,

crimefighter

[ChristianBale,

new.head()

₹	m	novie_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
new[' new.h	_	-	tags'].apply(lambda x: " ".join	(x))

∑			4:41.	****
_		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
	•	40000	The Deals Mainlet Diese	Following the death of District Attorney

send... [Following,

the death

[Action, Crime,

The Dark

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

```
vector = cv.fit_transform(new['tags']).toarray()
vector.shape
→ (937, 5000)
from sklearn.metrics.pairwise import cosine_similarity
similarity = cosine_similarity(vector)
similarity
→ array([[1.
                      , 0.08282364, 0.05437272, ..., 0.02602896, 0.09188458,
            0.
                      ],
            [0.08282364, 1.
                                , 0.06154575, ..., 0.02946278, 0.06933752,
            0.
                      ],
            [0.05437272, 0.06154575, 1. , ..., 0.02901294, 0.03413944,
            0.03711348],
            . . . ,
            [0.02602896, 0.02946278, 0.02901294, ..., 1.
                                                             , 0.13074409,
                   ],
            [0.09188458,\ 0.06933752,\ 0.03413944,\ \dots,\ 0.13074409,\ 1.
                ],
            0.
                      , 0.
                                  , 0.03711348, ..., 0.
                                                           , 0.
            [0.
            1.
                      ]])
new[new['title'] == 'The Lego Movie'].index[0]
<del>→</del> 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('Avatar')

→ Titan A.E.

    Ender's Game
     Battle: Los Angeles
    Independence Day
     Edge of Tomorrow
import pickle
pickle.dump(new,open('movie_list.pkl','wb'))
pickle.dump(similarity,open('similarity.pkl','wb'))
```