```python
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('/kaggle/input/tmdb-movie-metadata/tmdb_5000_movies.csv')
credits = pd.read_csv('/kaggle/input/tmdb-movie-metadata/tmdb_5000_credits.csv')
movies.head(2)
```
budget genres homepage id keywords original_language original_title overview popularity production_companies production_countries release_date revenue runtime spoken_languages status
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 para
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 Cari

```python
movies.shape
```
(4803, 20)
```python
credits.head()
```
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...
```python
movies = movies.merge(credits,on='title')
movies.head()
# budget
# homepage
# id
# original_language
# original_title
# popularity
# production_comapny
# production_countries
# release-date(not sure)
```
movie_id title genres overview cast crew
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,
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":
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,
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,... [{"
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":

```python
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, "nam... [{"id": 1463, "name": "culture clash"}, {"id":... [{"cast_id": 242,
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":
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 Bon
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, "cha
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

```python
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, "cha
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
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 Bon
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, "cha
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

```python
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, "cha
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
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 Bo
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, "cha
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

```python
import ast
ast.literal_eval('[{"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
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
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 Se
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, Mic
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 Colli

```python
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)
```
movie_id title overview genres keywords cast crew
1183 9889 Shallow Hal A shallow man falls in love with a 300 pound w... [Comedy, Romance] [overweight, beauty, hypnosis, overweight man,... [Gwyneth Paltrow, Jack Black, Jason Alexand
4215 15976 The Bubble The movie follows a group of young friends in ... [Drama, Romance] [gay] [Ohad Knoller, Yousef Sweid, Daniella Wircer] [Eytan Fox]
1571 22947 Up in the Air George Clooney plays the dry cynical character... [Drama, Romance] [suitcase, business, omaha, on the road, downs... [George Clooney, Vera Farmiga, Anna Kendr
205 58574 Sherlock Holmes: A Game of Shadows There is a new criminal mastermind at large (P... [Adventure, Action, Crime, Mystery] [detective inspector, steampunk, criminal mast... [R
352 10674 Mulan A tomboyish girl disguises herself as a young ... [Animation, Family, Adventure] [homeland, musical, training, daughter, cricke... [Eddie Murphy, Jackie Chan, Ming-Na

```python
def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ",""))
    return L1
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 genres keywords cast crew
0 19995 Avatar In the 22nd century, a paraplegic Marine is di... [Action, Adventure, Fantasy, ScienceFiction] [cultureclash, future, spacewar, spacecolony, ... [SamWorthington, ZoeSal
1 285 Pirates of the Caribbean: At World's End Captain Barbossa, long believed to be dead, ha... [Adventure, Fantasy, Action] [ocean, drugabuse, exoticisland, eastindiatrad... [JohnnyD
2 206647 Spectre A cryptic message from Bond's past sends him o... [Action, Adventure, Crime] [spy, basedonnovel, secretagent, sequel, mi6, ... [DanielCraig, ChristophWaltz, LéaSeydo
3 49026 The Dark Knight Rises Following the death of District Attorney Harve... [Action, Crime, Drama, Thriller] [dccomics, crimefighter, terrorist, secretiden... [ChristianBale, Mich
4 49529 John Carter John Carter is a war-weary, former military ca... [Action, Adventure, ScienceFiction] [basedonnovel, mars, medallion, spacetravel, p... [TaylorKitsch, LynnCollins,

```python
movies['overview'] = movies['overview'].apply(lambda x:x.split())
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
new = movies.drop(columns=['overview','genres','keywords','cast','crew'])
```

```python
#new.head()
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
from sklearn.metrics.pairwise import cosine_similarity
similarity = cosine_similarity(vector)
similarity
new[new['title'] == 'The Lego Movie'].index[0]
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('Gandhi')
Gandhi, My Father
The Wind That Shakes the Barley
A Passage to India
Guiana 1838
Ramanujan
import pickle
pickle.dump(new,open('movie_list.pkl','wb'))
pickle.dump(similarity,open('similarity.pkl','wb'))
```