Name: Adwait Purao
UID: 2021300101
DIV: B
BATCH: B2

```python
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import paudas as pd # data processing, CSV file I/0 (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

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

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & 1
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

```
/kaggle/input/tmdb-movie-metadata/tmdb_5000_movies.csv
/kaggle/input/tmdb-movie-metadata/tmdb_5000_credits.csv
```

```python
novles = pd.read_csv( '/kaggle/Input/tndb-novle-metadata/tmdb_5066_movies.csv ')
credits = pd.read_csv('/kaggle/input/tmdb-movie-metadata/tmdb_5000_credits.csv')
```

```python
movies.head()
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	populari
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...	150.437*
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...	139.082E
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	107.3767
3	250000000	[("id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	g t '84 "namd comics"), {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312.°
4	260000000	[("id": 28, "name": "Action"}, ("id": 12, "nam...	http:IImovies.disney.com/john-carter	49529	[{"id": 818, "name . "based on novel"), ("id":...	en	John Carter	John Carter is a war-weary, former military ca...	43.926c

```python
movies.shape
```

```
(4803, 20)
```

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

```
credits.head()
```

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

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

```
movies  =  movies.menge(cnedits,on='title')
```

```
movies.head()
# budget
# homepage
# Id
#  original_language
#   original_title
#   popularity
# production_comapny
# production_countries
#  release-date(not   sune)
```

	budget	**genres**	homepage	1d	keywords or1gJ
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...
	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": "•py"1. {"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"),

5 rows   29 columns

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

```
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", "...	[{"cr

```
import ast
```

| | 285 | Pirates of Caribbean: At WOrld's End | Barbossa, long believed to be dead, | [{"id": 12, "Adventure"), {"id": 14,"... | [{"id": 270, "name": "ocean"}, {"id": 726, "na. | [{"cast id": 4, "character": "Captain Jack Spa... | [{"cr 52fe4232c3a36847f80 |

```
de-r-  convert(text) •
  L — []
    fdr i in ast.literal_eval(text):
        L.append(i['name'])
    return L
```

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

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

	rtiov1e_1d	t1t1e	overv1ew	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", "...	[{"cre‹ 52fe48009251416c750a
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead,	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"), {"id": 726, "na...	[{"cast id": 4, "character": "Captain Jack Spa...	[{"cre‹ 52fe4232c3a36847f800l

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

	movie_id	t1t1e	overv1ew	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", "...	[{"crew 52fe48009251416c750a
1	285	Pirates of the Caribbean: At World's	Captain Barbossa, long believed to be	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island,	[{"cast_id": 4, "character": "Captain	[{"crew 52fe4232c3a36847f800

```
import ast
ast.literal_eval('[{"id": 28, "uame": "Action"}, {"id": 12, "uame": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"Id": 878, "name": "Sclenc
```

```
    [{'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.appeud(i['name'])
        counter+=1
    return L
```

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

	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...	[(" * "52fe48009251416c750
1	285	Pirates of the Caribbean: AtWorld's	Captain Barbossa, long believed to be	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island,	[Johnny Depp, Orlando Bloom, Keira	[(" ° "52fe4232c3a36847f80

```
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'] == 'Dinector':
            L.append(i['name])
    return L
```

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

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

	nov1e_1d	t1t1e	overv1ew	genres	keywords	**cast**	crew
3792	8337	They Live	Nada, a down-on-his-luck construction worker, ...	[Action, Horror, Science Fiction, Thriller]	[dystopia, social commentary, alien invasion,	[Roddy Piper, Keith David, Meg Foster]	[John Carpenter]
1070	64328	The Muppets	When Kermit the Frog and the Muppets learn tha...	[Comedy, Family, Music]	[musical, the muppets, robot, oil tycoon, stud...	[Amy Adams, Jason Segel, Emily Blunt]	[James Bobin]
			An American			[Paul Newman	

```
def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ",""))
    return L1
```

```
movies['cast'] = movies['cast'].apply(collapse)
movies['cnew'] = movies['cnew'].apply(collapse)
movies['genres'] = movies['genres'].apply(collapse)
movies['keywords'] = movies['keywords'].apply(collapse)
```

```
movies.head()
```

	movie_id	title	overview	genres	keyuiords	cast	
0	19995	Avatar	In the 22nd century, a paraplegic Marine is	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony,	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesC

```
movies['overview'] = movies['overvi4W!].apply(lambda x:x.split())"
```

```
movies['tags'] = nov&lT3@&99r@ @W.movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
```

			Captain				
1	285	Caribbean: the End	believed dead,	[Adventure, Fantasy, Action]	[ocea exoticisland eastindiatrad...	[JohnnyDepp, OrlandoBloom, jqiraKnightley]	[Gore\

```
new = novles.drop(c4Btrl/l¥Bf | overv1Vg 'genres 'Akeyords 'T'cast'; 'crew'l
#new.head()
```

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

	mov1e_1d	t1t1e	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...
	28s	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
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...

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

```
vector = cv.fit_tnansform(new['tags']).toarnay()
```

```
vector.shape
```

```
(4842, 5000)
```

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

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

```
similarity
```

```
array([[1.        , 0.09107651, 0.06071767, ..., 0.02492224, 0.02817181,
        0.        ],
       [0.09107651, 1.        , 0. 66451613,       0.02648136, 0.
        0.        ],
       [0.06071767, 0.06451613, 1.        ,       0.02648136, 0.
        0.        ],

       [0.02492224, 0.02648136, 0.02648136,       1.          0.07372098,
        0.04783649],
       [0.02817181, 0.        , 0.        ,       0.07372098, 1.
        0.05407381],
       [0.        , 0.        , 0.        ,       0.04783649, 0.05407381,
```

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

```
744
```

```
def recommend(movie):
    index = new[new['title'] == movie].iudex[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)
```

Double-click (or enter) to edit

```
reconnend( 'Avatar' )
```

```
Titan A. E .
Small  Soldiers
Independence Day
Ender ' s Gane
Aliens vs Predator: Requiem
```

```
import pickle
```

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