```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 pandas as pd # data processing, CSV file I/O (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 & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

```python
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [ ]:  !pip install kaggle
         ! mkdir ~/.kaggle
         ! cp kaggle.json ~/.kaggle/
         ! chmod 600 ~/.kaggle/kaggle.json
         !kaggle datasets download -d tmdb/tmdb-movie-metadata
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.10/dist-packages (1.5.15)
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.10/dist-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from kaggle) (2023.5.7)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.8.
2)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.27.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from kaggle) (4.65.0)
Requirement already satisfied: python-slugify in /usr/local/lib/python3.10/dist-packages (from kaggle) (8.0.
1)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from kaggle) (1.26.16)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from kaggle) (6.0.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->kaggle)
(0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.10/dist-packages (from python-sl
ugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from req
uests->kaggle) (2.0.12)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->kaggl
e) (3.4)
Downloading tmdb-movie-metadata.zip to /content
 90% 8.00M/8.89M [00:01<00:00, 12.6MB/s]
100% 8.89M/8.89M [00:01<00:00, 7.63MB/s]
```

```
In [ ]:  !!unzip /content/tmdb-movie-metadata.zip
```

```
Archive:  /content/tmdb-movie-metadata.zip
  inflating: tmdb_5000_credits.csv
  inflating: tmdb_5000_movies.csv
```

```
In [ ]:  movies = pd.read_csv('/content/tmdb_5000_movies.csv')
         credits = pd.read_csv('/content/tmdb_5000_credits.csv')
```

```
In [ ]:   movies.head(2)
```

Out[ ]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popula
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...	150.4375
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...	139.0826

```
In [ ]:   movies.shape
```

Out[ ]:   (4803, 20)

```
In [ ]:   credits.head()
```

Out[ ]:

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

```
In [ ]:  movies.head()
         # budget
         # homepage
         # id
         # original_language
         # original_title
         # popularity
         # production_comapny
         # production_countries
         # release-date(not sure)
```

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

```
In [ ]:  movies.head()
```

Out[ ]:

	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, "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, "...	[{"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...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"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...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
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...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [ ]:   movies.shape

Out[ ]:   (4809, 7)


In [ ]:   import ast


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


In [ ]:   movies.dropna(inplace=True)


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

Out[ ]:

	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...
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...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

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

Out[ ]:

	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...
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...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```python
In [ ]:   import ast
          ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantas
          y"}, {"id": 878, "name": "Science Fiction"}]')
```

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

```
In [ ]:  #taking only 1st 3 cast
         def convert3(text):
             L = []
             counter = 0
             for i in ast.literal_eval(text):
                 if counter < 3:
                     L.append(i['name'])
                 counter+=1
             return L
```

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

Out[ ]:

	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...
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,...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [ ]:  movies['cast'][0]
```

```
Out[ ]:  ['Sam Worthington', 'Zoe Saldana', 'Sigourney Weaver']
```

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

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

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

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

Out[ ]:

	movie_id	title	overview	genres	keywords	cast	crew
1922	9787	Lords of Dogtown	The film follows the surf and skateboarding tr...	[Drama]	[skateboarding, success, independent film, wom...	[Emile Hirsch, Victor Rasuk, Rebecca De Mornay]	[Catherine Hardwicke]
1289	110415	Snowpiercer	In a future where a failed global-warming expe...	[Action, Science Fiction, Drama]	[father son relationship, child labour, brothe...	[Chris Evans, Song Kang-ho, Ed Harris]	[Bong Joon-ho]
4759	193603	Hayride	A college student returning home for Halloween...	[Thriller, Horror]	[haunted house, slasher]	[Richard Tyson, Sherri Eakin, Jeremy Ivy]	[Terron R. Parsons]
1402	609	Poltergeist	Steve Freeling lives with his wife, Diane, and...	[Horror]	[medium, ghostbuster, poltergeist, haunted hou...	[Craig T. Nelson, JoBeth Williams, Beatrice St...	[Tobe Hooper]
1196	11362	The Count of Monte Cristo	Edmond Dantés's life and plans to marry the be...	[Action, Adventure, Drama, Thriller]	[loss of lover, lover (female), ex-lover, tort...	[Jim Caviezel, Guy Pearce, Richard Harris]	[Kevin Reynolds]

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

remove spaces between words inbetween commas

```
In [ ]:  #converting to list
         movies['cast'] = movies['cast'].apply(collapse)
         movies['crew'] = movies['crew'].apply(collapse)
         movies['genres'] = movies['genres'].apply(collapse)
         movies['keywords'] = movies['keywords'].apply(collapse)
```

```
In [ ]:  movies.head(1)
```

Out[ ]:

	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, ZoeSaldana, SigourneyWeaver]	[JamesCameron]

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

```
In [ ]:  movies.head(1)
```

Out[ ]:

	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]

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

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

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

Out[ ]:

	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...

```
In [ ]: new['tags'][0]
```

Out[ ]: '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. Action Adventure Fantasy ScienceFiction cultureclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCameron'

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

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

```
In [ ]: vector.shape
```

Out[ ]: (4806, 5000)

```
In [ ]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [ ]: similarity = cosine_similarity(vector)
```

```
In [ ]: similarity
```

```
Out[ ]: array([[1.        , 0.08964215, 0.06071767, ..., 0.02519763, 0.0277885 ,
                 0.        ],
                [0.08964215, 1.        , 0.06350006, ..., 0.02635231, 0.        ,
                 0.        ],
                [0.06071767, 0.06350006, 1.        , ..., 0.02677398, 0.        ,
                 0.        ],
                ...,
                [0.02519763, 0.02635231, 0.02677398, ..., 1.        , 0.07352146,
                 0.04774099],
                [0.0277885 , 0.        , 0.        , ..., 0.07352146, 1.        ,
                 0.05264981],
                [0.        , 0.        , 0.        , ..., 0.04774099, 0.05264981,
                 1.        ]])
```

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

```
Out[ ]: 744
```

```
In [ ]: similarity[744]
```

```
Out[ ]: array([0.06761234, 0.07071068, 0.03592106, ..., 0.02981424, 0.        ,
                 0.        ])
```

```python
In [ ]: def recommend(movie):
            index = new[new['title'] == movie].index[0]
            distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda x: x[1]) #x[1] specifies
        sorting wrt column 2 i.e similarity and not on index
            for i in distances[1:6]:
                print(new.iloc[i[0]].title)
```

```
In [ ]: recommend("Pirates of the Caribbean: At World's End")

        Pirates of the Caribbean: Dead Man's Chest
        Pirates of the Caribbean: The Curse of the Black Pearl
        Pirates of the Caribbean: On Stranger Tides
        20,000 Leagues Under the Sea
        Puss in Boots
```

```
In [ ]: import pickle
```

```
In [ ]: pickle.dump(new,open('movie_list.pkl','wb'))
        pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]:
```