In [1]:
```
!pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\admin\anaconda3\lib\site-packages (3.8.1)
Requirement already satisfied: click in c:\users\admin\anaconda3\lib\site-packages (from nltk) (8.0.4)
Requirement already satisfied: joblib in c:\users\admin\anaconda3\lib\site-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in c:\users\admin\anaconda3\lib\site-packages (from nlt
k) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\admin\anaconda3\lib\site-packages (from nltk) (4.65.0)
Requirement already satisfied: colorama in c:\users\admin\anaconda3\lib\site-packages (from click->nlt
k) (0.4.6)
```

In [5]:
```
import re
import ast
import pickle

import numpy as np
import pandas as pd
import seaborn as sns

import nltk
from nltk.stem.porter import PorterStemmer
from nltk.corpus import stopwords

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity

import warnings
warnings.filterwarnings('ignore')
```

In [ ]:
```

```

In [6]:
```
movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

In [ ]:

In [7]:
```
movies.head(3)
```

Out[7]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview
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...
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...
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...

In [8]:
```
credits.head()
```

Out[8]:

	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 [10]:
```
movies.shape, credits.shape
```

Out[10]:
```
((4803, 20), (4803, 4))
```

In [11]:
```
movies.columns
```

Out[11]:
```
Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
       'original_title', 'overview', 'popularity', 'production_companies',
       'production_countries', 'release_date', 'revenue', 'runtime',
       'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
       'vote_count'],
      dtype='object')
```

In [12]:
```
credits.columns
```

Out[12]:
```
Index(['movie_id', 'title', 'cast', 'crew'], dtype='object')
```

In [13]:
```
# Merge dataframes

movies = movies.merge(credits, on='title')
movies.head(3)
```

Out[13]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview
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...
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...
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...

In [ ]:
```
```

Choose the relevant features for movie recommendation

- movie_id
- title
- overview
- genres
- keywords
- cast

- crew

In [25]:
```
movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']][:2]
```

Out[25]:

	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...

In [ ]:
```
```

In [26]:
```
df = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]
df.head(3)
```

Out[26]:

	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...

In [27]:
```
df.shape
```

Out[27]:
```
(4809, 7)
```

In [28]:
```
df.isnull().sum()
```

Out[28]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [31]:
```
df['overview'].fillna(' ', inplace=True)
```

In [32]:
```
df.isnull().sum()
```

Out[32]:
```
movie_id    0
title       0
overview    0
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [ ]:
```

```

Final Goal - movie_id + title + tags

In [33]:
```
df.head(2)
```

Out[33]:

	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...

In [ ]:
```
```

Let's clean genres feature firstly

In [34]:
```
df['genres']
```

Out[34]:
```
0       [{"id": 28, "name": "Action"}, {"id": 12, "nam...
1       [{"id": 12, "name": "Adventure"}, {"id": 14, "...
2       [{"id": 28, "name": "Action"}, {"id": 12, "nam...
3       [{"id": 28, "name": "Action"}, {"id": 80, "nam...
4       [{"id": 28, "name": "Action"}, {"id": 12, "nam...
                              ...
4804    [{"id": 28, "name": "Action"}, {"id": 80, "nam...
4805    [{"id": 35, "name": "Comedy"}, {"id": 10749, "...
4806    [{"id": 35, "name": "Comedy"}, {"id": 18, "nam...
4807                                                   []
4808                    [{"id": 99, "name": "Documentary"}]
Name: genres, Length: 4809, dtype: object
```

In [35]:
```
df['genres'][0]
```

Out[35]:
```
'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id":
878, "name": "Science Fiction"}]'
```

In [36]:
```
type(df['genres'][0])
```

Out[36]:
```
str
```

In [38]:
```
ast.literal_eval(df['genres'][0])
```

Out[38]:
```
[{'id': 28, 'name': 'Action'},
 {'id': 12, 'name': 'Adventure'},
 {'id': 14, 'name': 'Fantasy'},
 {'id': 878, 'name': 'Science Fiction'}]
```

In [41]:
```
for i in ast.literal_eval(df['genres'][0]):
    print(i['name'])
```

```
Action
Adventure
Fantasy
Science Fiction
```

In [44]:
```
def fetch_genres(text):

    l = []

    for i in ast.literal_eval(text):
        l.append(i['name'])

    return l
```

In [48]:
```
fetch_genres(df['genres'][0])
```

Out[48]:
```
['Action', 'Adventure', 'Fantasy', 'Science Fiction']
```

In [49]:
```
df['genres'].apply(fetch_genres)
```

Out[49]:
```
0       [Action, Adventure, Fantasy, Science Fiction]
1                     [Adventure, Fantasy, Action]
2                       [Action, Adventure, Crime]
3                   [Action, Crime, Drama, Thriller]
4               [Action, Adventure, Science Fiction]
                           ...
4804                      [Action, Crime, Thriller]
4805                             [Comedy, Romance]
4806            [Comedy, Drama, Romance, TV Movie]
4807                                            []
4808                                  [Documentary]
Name: genres, Length: 4809, dtype: object
```

In [ ]:
```

```

In [50]:
```
df[:2]
```

Out[50]:

	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...

In [51]:
```
df['genres'] = df['genres'].apply(fetch_genres)
```

In [54]:
```
df[:3]
```

Out[54]:

	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...

In [ ]:
```

```

Keywords

In [55]:
```
df['keywords']
```

Out[55]:
```
0       [{"id": 1463, "name": "culture clash"}, {"id":...
1       [{"id": 270, "name": "ocean"}, {"id": 726, "na...
2       [{"id": 470, "name": "spy"}, {"id": 818, "name...
3       [{"id": 849, "name": "dc comics"}, {"id": 853,...
4       [{"id": 818, "name": "based on novel"}, {"id":...
                              ...
4804    [{"id": 5616, "name": "united states\u2013mexi...
4805                                                   []
4806    [{"id": 248, "name": "date"}, {"id": 699, "nam...
4807                                                   []
4808    [{"id": 1523, "name": "obsession"}, {"id": 224...
Name: keywords, Length: 4809, dtype: object
```

In [56]:
```
df['keywords'][0]
```

Out[56]:
```
'[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space wa
r"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "society"}, {"id": 3801, "name": "space
travel"}, {"id": 9685, "name": "futuristic"}, {"id": 9840, "name": "romance"}, {"id": 9882, "name": "sp
ace"}, {"id": 9951, "name": "alien"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien plan
et"}, {"id": 10987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldier"},
{"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165431, "name": "anti wa
r"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "name": "mind and soul"}, {"id": 209714,
"name": "3d"}]'
```

In [ ]:
```
```

In [57]:
```
def fetch_keywords(text):

    l = []

    for i in ast.literal_eval(text):
        l.append(i['name'])

    return l
```

In [58]:
```
fetch_keywords(df['keywords'][0])
```

Out[58]:
```
['culture clash',
 'future',
 'space war',
 'space colony',
 'society',
 'space travel',
 'futuristic',
 'romance',
 'space',
 'alien',
 'tribe',
 'alien planet',
 'cgi',
 'marine',
 'soldier',
 'battle',
 'love affair',
 'anti war',
 'power relations',
 'mind and soul',
 '3d']
```

In [59]:
```
df['keywords'].apply(fetch_keywords)
```

Out[59]:
```
0       [culture clash, future, space war, space colon...
1       [ocean, drug abuse, exotic island, east india ...
2       [spy, based on novel, secret agent, sequel, mi...
3       [dc comics, crime fighter, terrorist, secret i...
4       [based on novel, mars, medallion, space travel...
                              ...
4804    [united states-mexico barrier, legs, arms, pap...
4805                                                   []
4806    [date, love at first sight, narration, investi...
4807                                                   []
4808              [obsession, camcorder, crush, dream girl]
Name: keywords, Length: 4809, dtype: object
```

In [60]:
```
df['keywords'] = df['keywords'].apply(fetch_keywords)
```

In [61]:
```
df.head(3)
```

Out[61]:

	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...

In [ ]:
```

```

Cast

In [62]:
```
df['cast']
```

Out[62]:
```
0       [{"cast_id": 242, "character": "Jake Sully", "...
1       [{"cast_id": 4, "character": "Captain Jack Spa...
2       [{"cast_id": 1, "character": "James Bond", "cr...
3       [{"cast_id": 2, "character": "Bruce Wayne / Ba...
4       [{"cast_id": 5, "character": "John Carter", "c...
                              ...
4804    [{"cast_id": 1, "character": "El Mariachi", "c...
4805    [{"cast_id": 1, "character": "Buzzy", "credit_...
4806    [{"cast_id": 8, "character": "Oliver O\u2019To...
4807    [{"cast_id": 3, "character": "Sam", "credit_id...
4808    [{"cast_id": 3, "character": "Herself", "credi...
Name: cast, Length: 4809, dtype: object
```

In [63]:
```
df['cast'][0]
```

Out[63]:
```
'[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2,
"id": 65731, "name": "Sam Worthington", "order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_
id": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana", "order": 1}, {"cast
_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52fe48009251416c750aca39", "gender": 1,
"id": 10205, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quaritch",
"credit_id": "52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang", "order":
3}, {"cast_id": 5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d3", "gender":
1, "id": 17647, "name": "Michelle Rodriguez", "order": 4}, {"cast_id": 8, "character": "Selfridge",
"credit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovanni Ribisi", "orde
r": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id": "52fe48009251416c750ac9dd", "gende
r": 2, "id": 59231, "name": "Joel David Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "cr
edit_id": "52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder", "order": 7},
{"cast_id": 11, "character": "Eytukan", "credit_id": "52fe48009251416c750ac9ed", "gender": 2, "id":
15853, "name": "Wes Studi", "order": 8}, {"cast_id": 10, "character": "Tsu\'Tey", "credit_id": "52fe
48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso", "order": 9}, {"cast_id": 12,
"character": "Dr. Max Patel", "credit_id": "52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "na
me": "Dileep Rao", "order": 10}, {"cast_id": 13, "character": "Lyle Wainfleet", "credit_id": "52fe48
009251416c750ac9f5", "gender": 2, "id": 98215, "name": "Matt Gerald", "order": 11}, {"cast_id": 32,
"character": "Private Fike", "credit_id": "52fe48009251416c750aca5b", "gender": 2, "id": 154153, "na
```

In [ ]:

```
```

In [68]:

```
def fetch_cast(text):

    l = []
    counter = 0

    for i in ast.literal_eval(text):
        if counter != 3:
            l.append(i['name'])
            counter += 1
        else:
            break

    return l
```

In [69]:

```
fetch_cast(df['cast'][0])
```

Out[69]:

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

In [70]:

```
df['cast'].apply(fetch_cast)
```

Out[70]:

```
0        [Sam Worthington, Zoe Saldana, Sigourney Weaver]
1           [Johnny Depp, Orlando Bloom, Keira Knightley]
2            [Daniel Craig, Christoph Waltz, Léa Seydoux]
3            [Christian Bale, Michael Caine, Gary Oldman]
4          [Taylor Kitsch, Lynn Collins, Samantha Morton]
                              ...
4804    [Carlos Gallardo, Jaime de Hoyos, Peter Marqua...
4805         [Edward Burns, Kerry Bishé, Marsha Dietlein]
4806           [Eric Mabius, Kristin Booth, Crystal Lowe]
4807            [Daniel Henney, Eliza Coupe, Bill Paxton]
4808    [Drew Barrymore, Brian Herzlinger, Corey Feldman]
Name: cast, Length: 4809, dtype: object
```

In [71]:
```
df['cast'] = df['cast'].apply(fetch_cast)
```

In [72]:
```
df[:3]
```

Out[72]:

	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 Weaver]	[{"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]	[{"credit_id": "54805967c3a36829b5002c41", "de...

In [ ]:
```

```

Crew

In [73]:
```
df['crew']
```

Out[73]:
```
0       [{"credit_id": "52fe48009251416c750aca23", "de...
1       [{"credit_id": "52fe4232c3a36847f800b579", "de...
2       [{"credit_id": "54805967c3a36829b5002c41", "de...
3       [{"credit_id": "52fe4781c3a36847f81398c3", "de...
4       [{"credit_id": "52fe479ac3a36847f813eaa3", "de...
                              ...
4804    [{"credit_id": "52fe44eec3a36847f80b280b", "de...
4805    [{"credit_id": "52fe487dc3a368484e0fb013", "de...
4806    [{"credit_id": "52fe4df3c3a36847f8275ecf", "de...
4807    [{"credit_id": "52fe4ad9c3a368484e16a36b", "de...
4808    [{"credit_id": "58ce021b9251415a390165d9", "de...
Name: crew, Length: 4809, dtype: object
```

In [74]:
```
df['crew'][0]
```

Out[74]:
```
'[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job":
"Editor", "name": "Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Ar
t", "gender": 2, "id": 496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491
c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "job": "Sound Designer", "nam
e": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gender":
0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "539c4a
4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name":
"Mali Finn"}, {"credit_id": "5544ee3b925141499f0008fc", "department": "Sound", "gender": 2, "id": 17
29, "job": "Original Music Composer", "name": "James Horner"}, {"credit_id": "52fe48009251416c750ac9
c3", "department": "Directing", "gender": 2, "id": 2710, "job": "Director", "name": "James Camero
n"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "gender": 2, "id": 2710, "jo
b": "Writer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "department": "Edi
ting", "gender": 2, "id": 2710, "job": "Editor", "name": "James Cameron"}, {"credit_id": "52fe480092
51416c750aca29", "department": "Production", "gender": 2, "id": 2710, "job": "Producer", "name": "Ja
mes Cameron"}, {"credit_id": "52fe48009251416c750aca3f", "department": "Writing", "gender": 2, "id":
2710, "job": "Screenplay", "name": "James Cameron"}, {"credit_id": "539c4987c3a36810ba0021a4", "depa
rtment": "Art", "gender": 2, "id": 7236, "job": "Art Direction", "name": "Andrew Menzies"}, {"credit
_id": "549598c3c3a3686ae9004383", "department": "Visual Effects", "gender": 0, "id": 6690, "job": "V
isual Effects Producer", "name": "Jill Brooks"}, {"credit_id": "52fe48009251416c750aca4b", "departme
nt": "Production", "gender": 1, "id": 6347, "job": "Casting", "name": "Margery Simkin"}, {"credit_i
```

In [ ]:
```
```

In [77]:
```
def fetch_director(text):

    l = []

    for i in ast.literal_eval(text):
        if i['job'] == 'Director':
            l.append(i['name'])

    return l
```

In [78]:
```
fetch_director(df['crew'][0])
```

Out[78]:
```
['James Cameron']
```

In [79]:
```
df['crew']df['crew'].apply(fetch_director)
```

Out[79]:
```
0                                [James Cameron]
1                               [Gore Verbinski]
2                                    [Sam Mendes]
3                              [Christopher Nolan]
4                                 [Andrew Stanton]
                              ...
4804                             [Robert Rodriguez]
4805                                 [Edward Burns]
4806                                  [Scott Smith]
4807                                  [Daniel Hsia]
4808    [Brian Herzlinger, Jon Gunn, Brett Winn]
Name: crew, Length: 4809, dtype: object
```

In [81]:
```
df['crew'] = df['crew'].apply(fetch_director)
```

In [82]:
```
df.head()
```

Out[82]:

	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 Weaver]	[James Cameron]
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]	[Gore Verbinski]
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]	[Sam Mendes]
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]	[Christopher Nolan]
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]	[Andrew Stanton]

In [ ]:
```

```

Overview

In [83]:
```
df['overview']
```

Out[83]:
```
0       In the 22nd century, a paraplegic Marine is di...
1       Captain Barbossa, long believed to be dead, ha...
2       A cryptic message from Bond’s past sends him o...
3       Following the death of District Attorney Harve...
4       John Carter is a war-weary, former military ca...
                              ...
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: overview, Length: 4809, dtype: object
```

In [84]:
```
df['overview'][0]
```

Out[84]:
```
'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but be
comes torn between following orders and protecting an alien civilization.'
```

In [85]:
```
'i am a boy'
```

Out[85]:
```
'i am a boy'
```

In [86]:
```
'i am a boy'.split()
```

Out[86]:
```
['i', 'am', 'a', 'boy']
```

In [87]:
```
df['overview'][0].split()
```

Out[87]:
```
['In',
 'the',
 '22nd',
 'century,',
 'a',
 'paraplegic',
 'Marine',
 'is',
 'dispatched',
 'to',
 'the',
 'moon',
 'Pandora',
 'on',
 'a',
 'unique',
 'mission,',
 'but',
 'becomes',
```

In [ ]:
```

```

In [88]:
```
df['overview'].apply(lambda x: x.split())
```

Out[88]:
```
0       [In, the, 22nd, century,, a, paraplegic, Marin...
1       [Captain, Barbossa,, long, believed, to, be, d...
2       [A, cryptic, message, from, Bond’s, past, send...
3       [Following, the, death, of, District, Attorney...
4       [John, Carter, is, a, war-weary,, former, mili...
                              ...
4804    [El, Mariachi, just, wants, to, play, his, gui...
4805    [A, newlywed, couple's, honeymoon, is, upended...
4806    ["Signed,, Sealed,, Delivered", introduces, a,...
4807    [When, ambitious, New, York, attorney, Sam, is...
4808    [Ever, since, the, second, grade, when, he, fi...
Name: overview, Length: 4809, dtype: object
```

In [89]:
```
df['overview'] = df['overview'].apply(lambda x: x.split())
```

In [90]:
```
df.head()
```

Out[90]:

	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]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

In [91]:
```
[3,4] + [5]
```

Out[91]:
```
[3, 4, 5]
```

In [92]:
```
['a', 'b'] + ['c']
```

Out[92]:
```
['a', 'b', 'c']
```

In [93]:
```
df['overview'] + df['genres'] + df['keywords'] + df['cast'] + df['crew']
```

Out[93]:
```
0       [In, the, 22nd, century,, a, paraplegic, Marin...
1       [Captain, Barbossa,, long, believed, to, be, d...
2       [A, cryptic, message, from, Bond’s, past, send...
3       [Following, the, death, of, District, Attorney...
4       [John, Carter, is, a, war-weary,, former, mili...
                              ...
4804    [El, Mariachi, just, wants, to, play, his, gui...
4805    [A, newlywed, couple's, honeymoon, is, upended...
4806    ["Signed,, Sealed,, Delivered", introduces, a,...
4807    [When, ambitious, New, York, attorney, Sam, is...
4808    [Ever, since, the, second, grade, when, he, fi...
Length: 4809, dtype: object
```

In [97]:
```
print((df['overview'] + df['genres'] + df['keywords'] + df['cast'] + df['crew'])[0])
```

```
['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', 'Science Fictio
n', 'culture clash', 'future', 'space war', 'space colony', 'society', 'space travel', 'futuristic', 'r
omance', 'space', 'alien', 'tribe', 'alien planet', 'cgi', 'marine', 'soldier', 'battle', 'love affai
r', 'anti war', 'power relations', 'mind and soul', '3d', 'Sam Worthington', 'Zoe Saldana', 'Sigourney
Weaver', 'James Cameron']
```

In [ ]:
```
```

In [98]:
```
df['tags'] = df['overview'] + df['genres'] + df['keywords'] + df['cast'] + df['crew']
```

In [99]:
```
df.head()
```

Out[99]:

	movie_id	title	overview	genres	keywords	cast	crew	tags
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]	[In, the, 22nd, century,, a, paraplegic, Marin...
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]	[Captain, Barbossa,, long, believed, to, be, d...
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]	[A, cryptic, message, from, Bond’s, past, send...
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]	[Following, the, death, of, District, Attorney...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]	[John, Carter, is, a, war-weary,, former, mili...

In [ ]:
```

```

Final Dataframe

In [100]:
```
df = df[['movie_id', 'title', 'tags']]
df
```

Out[100]:

	movie_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...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...
...	...	...	...
4804	9367	El Mariachi	[El, Mariachi, just, wants, to, play, his, gui...
4805	72766	Newlyweds	[A, newlywed, couple's, honeymoon, is, upended...
4806	231617	Signed, Sealed, Delivered	["Signed,, Sealed,, Delivered", introduces, a,...
4807	126186	Shanghai Calling	[When, ambitious, New, York, attorney, Sam, is...
4808	25975	My Date with Drew	[Ever, since, the, second, grade, when, he, fi...

4809 rows × 3 columns

In [ ]:
```

```

In [101]:
```
print(df['tags'][0])
```

```
['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', 'Science Fictio
n', 'culture clash', 'future', 'space war', 'space colony', 'society', 'space travel', 'futuristic', 'r
omance', 'space', 'alien', 'tribe', 'alien planet', 'cgi', 'marine', 'soldier', 'battle', 'love affai
r', 'anti war', 'power relations', 'mind and soul', '3d', 'Sam Worthington', 'Zoe Saldana', 'Sigourney
Weaver', 'James Cameron']
```

In [102]:
```
'i am a boy'
```

Out[102]:
```
'i am a boy'
```

In [103]:
```
'i am a boy'.replace(' ', '')
```

Out[103]:
```
'iamaboy'
```

In [104]:
```
df['tags'].apply(lambda x: [i.replace(' ', '') for i in x])
```

Out[104]:
```
0       [In, the, 22nd, century,, a, paraplegic, Marin...
1       [Captain, Barbossa,, long, believed, to, be, d...
2       [A, cryptic, message, from, Bond’s, past, send...
3       [Following, the, death, of, District, Attorney...
4       [John, Carter, is, a, war-weary,, former, mili...
                              ...
4804    [El, Mariachi, just, wants, to, play, his, gui...
4805    [A, newlywed, couple's, honeymoon, is, upended...
4806    ["Signed,, Sealed,, Delivered", introduces, a,...
4807    [When, ambitious, New, York, attorney, Sam, is...
4808    [Ever, since, the, second, grade, when, he, fi...
Name: tags, Length: 4809, dtype: object
```

In [105]:
```
df['tags'] = df['tags'].apply(lambda x: [i.replace(' ', '') for i in x])
```

In [106]:
```
print(df['tags'][0])
```

```
['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', 'ScienceFictio
n', 'cultureclash', 'future', 'spacewar', 'spacecolony', 'society', 'spacetravel', 'futuristic', 'roman
ce', 'space', 'alien', 'tribe', 'alienplanet', 'cgi', 'marine', 'soldier', 'battle', 'loveaffair', 'ant
iwar', 'powerrelations', 'mindandsoul', '3d', 'SamWorthington', 'ZoeSaldana', 'SigourneyWeaver', 'James
Cameron']
```

In [ ]:

```
```

In [107]:

```
df.head()
```

Out[107]:

	movie_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...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...

In [108]:

```
['i', 'am', 'a', 'boy']
```

Out[108]:

```
['i', 'am', 'a', 'boy']
```

In [111]:

```
" ".join(['i', 'am', 'a', 'boy'])
```

Out[111]:

```
'i am a boy'
```

In [ ]:

```
```

In [112]:
```
df['tags'].apply(lambda x: " ".join(x))
```

Out[112]:
```
0       In the 22nd century, a paraplegic Marine is di...
1       Captain Barbossa, long believed to be dead, ha...
2       A cryptic message from Bond’s past sends him o...
3       Following the death of District Attorney Harve...
4       John Carter is a war-weary, former military ca...
                              ...
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: tags, Length: 4809, dtype: object
```

In [113]:
```
df['tags'] = df['tags'].apply(lambda x: " ".join(x))
```

In [114]:
```
df
```

Out[114]:

	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...
...	...	...	...
4804	9367	El Mariachi	El Mariachi just wants to play his guitar and ...
4805	72766	Newlyweds	A newlywed couple's honeymoon is upended by th...
4806	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...
4807	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...
4808	25975	My Date with Drew	Ever since the second grade when he first saw ...

4809 rows × 3 columns

In [ ]:

NLP concepts to preprocess text data

- Lower Case
- Tokenization
- Stemming
- Stopwords Removal

In [115]:
```
df.head()
```

Out[115]:

	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 [116]:
```
'SOURAV'.lower()
```

Out[116]:
```
'sourav'
```

In [117]:
```
['love', 'lover', 'lovers', 'loving', 'lovely']
```

Out[117]:
```
['love', 'lover', 'lovers', 'loving', 'lovely']
```

In [119]:
```
ps = PorterStemmer()

for i in ['love', 'lover', 'lovers', 'loving', 'lovely']:
    print(f'{i} ---> {ps.stem(i)}')
```

```
love ---> love
lover ---> lover
lovers ---> lover
loving ---> love
lovely ---> love
```

In [ ]:
```

```

Text Preprocessing

In [133]:
```
ps = PorterStemmer()

def text_preprocess(text):

    new_text = []

    for i in text.split():
        lower = i.lower()
        new_text.append(ps.stem(lower))

    return " ".join(new_text)
```

In [134]:
```
text_preprocess(df['tags'][0])
```

Out[134]:
```
'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom to
rn between follow order and protect an alien civilization. action adventur fantasi sciencefict culturec
lash futur spacewar spacecoloni societi spacetravel futurist romanc space alien tribe alienplanet cgi m
arin soldier battl loveaffair antiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav j
amescameron'
```

In [120]:
```
df['tags'][0]
```

Out[120]:
```
'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but be
comes torn between following orders and protecting an alien civilization. Action Adventure Fantasy Scie
nceFiction cultureclash future spacewar spacecolony society spacetravel futuristic romance space alien
tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthin
gton ZoeSaldana SigourneyWeaver JamesCameron'
```

In [ ]:
```

```

In [135]:
```
df['tags'].apply(text_preprocess)
```

Out[135]:
```
0       in the 22nd century, a parapleg marin is dispa...
1       captain barbossa, long believ to be dead, ha c...
2       a cryptic messag from bond’ past send him on a...
3       follow the death of district attorney harvey d...
4       john carter is a war-weary, former militari ca...
                              ...
4804    el mariachi just want to play hi guitar and ca...
4805    a newlyw couple' honeymoon is upend by the arr...
4806    "signed, sealed, delivered" introduc a dedic q...
4807    when ambiti new york attorney sam is sent to s...
4808    ever sinc the second grade when he first saw h...
Name: tags, Length: 4809, dtype: object
```

In [136]:
```
df['tags'] = df['tags'].apply(text_preprocess)
```

In [137]:
```
df
```

Out[137]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond’ past send him on a...
3	49026	The Dark Knight Rises	follow the death of district attorney harvey d...
4	49529	John Carter	john carter is a war-weary, former militari ca...
...	...	...	...
4804	9367	El Mariachi	el mariachi just want to play hi guitar and ca...
4805	72766	Newlyweds	a newlyw couple' honeymoon is upend by the arr...
4806	231617	Signed, Sealed, Delivered	"signed, sealed, delivered" introduc a dedic q...
4807	126186	Shanghai Calling	when ambiti new york attorney sam is sent to s...
4808	25975	My Date with Drew	ever sinc the second grade when he first saw h...

4809 rows × 3 columns

In [ ]:
```
```

In [141]:
```
df['tags'][1205]
```

Out[141]:
```
"district attorney tom logan is set for higher office, at least until he becom involv with defenc lawye
r laura kelli and her unpredict client chelsea deardon. it seem the least of chelsea' crime is the thef
t of a veri valuabl painting, but as the women persuad logan to investig further and to cut some offici
corners, a much more sinist scenario start to emerge. comedi crime drama romanc thriller courtcas clien
t lawyer courtroom robertredford debrawing darylhannah ivanreitman"
```

In [ ]:
```
```

BOW method to encode your text data

In [142]:
```
cv = CountVectorizer(max_features=5000, stop_words='english')
cv
```

Out[142]:
```
CountVectorizer(max_features=5000, stop_words='english')
```
(https://scikit-learn.org/1.6/modules/generated/sklearn.feature_extraction.t

In [143]:
```
cv.fit_transform(df['tags'])
```

Out[143]:
```
<4809x5000 sparse matrix of type '<class 'numpy.int64'>'
	with 145438 stored elements in Compressed Sparse Row format>
```

In [146]:
```
vectors = cv.fit_transform(df['tags']).toarray()
vectors
```

Out[146]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [147]:
```
vectors.shape
```

Out[147]:
```
(4809, 5000)
```

In [ ]:

Cosine Similarity

In [148]:
```
cosine_similarity(vectors)
```

Out[148]:
```
array([[1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
        0.        ],
       [0.08346223, 1.        , 0.06063391, ..., 0.02378257, 0.        ,
        0.02615329],
       [0.0860309 , 0.06063391, 1.        , ..., 0.02451452, 0.        ,
        0.        ],
       ...,
       [0.04499213, 0.02378257, 0.02451452, ..., 1.        , 0.03962144,
        0.04229549],
       [0.        , 0.        , 0.        , ..., 0.03962144, 1.        ,
        0.08714204],
       [0.        , 0.02615329, 0.        , ..., 0.04229549, 0.08714204,
        1.        ]])
```

In [152]:
```
similarity = cosine_similarity(vectors)
similarity
```

Out[152]:
```
array([[1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
        0.        ],
       [0.08346223, 1.        , 0.06063391, ..., 0.02378257, 0.        ,
        0.02615329],
       [0.0860309 , 0.06063391, 1.        , ..., 0.02451452, 0.        ,
        0.        ],
       ...,
       [0.04499213, 0.02378257, 0.02451452, ..., 1.        , 0.03962144,
        0.04229549],
       [0.        , 0.        , 0.        , ..., 0.03962144, 1.        ,
        0.08714204],
       [0.        , 0.02615329, 0.        , ..., 0.04229549, 0.08714204,
        1.        ]])
```

In [150]:
```
distance.shape
```

Out[150]:
```
(4809, 4809)
```

In [151]:
```
df[:5]
```

Out[151]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond’ past send him on a...
3	49026	The Dark Knight Rises	follow the death of district attorney harvey d...
4	49529	John Carter	john carter is a war-weary, former militari ca...

In [154]:
```
distance = similarity[0]
distance
```

Out[154]:
```
array([1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
       0.        ])
```

In [156]:
```
sorted(similarity[0], reverse=True)
```

Out[156]:
```
[1.0000000000000002,
 0.28676966733820225,
 0.26901379342448517,
 0.2605130246476754,
 0.255608593705383,
 0.25038669783359574,
 0.24511108480187255,
 0.24455799402225925,
 0.23179316248638276,
 0.23174488732966073,
 0.2278389747471728,
 0.2252817784447915,
 0.22269966704152225,
 0.21853668936906193,
 0.21239769762143662,
 0.2108663315950723,
 0.2105263157894737,
 0.20602141085758227,
 0.20443988269091456,
```

In [ ]:

In [159]:
```
list(enumerate(similarity[0]))
```

Out[159]:
```
[(0, 1.0000000000000002),
 (1, 0.08346223261119858),
 (2, 0.08603090020146065),
 (3, 0.0734718358370645),
 (4, 0.19134594929397597),
 (5, 0.10838874619051501),
 (6, 0.04024218182927669),
 (7, 0.14673479641335554),
 (8, 0.05923488777590923),
 (9, 0.0967301666813349),
 (10, 0.10259783520851541),
 (11, 0.09464970485606021),
 (12, 0.09037128496931669),
 (13, 0.04499212706658476),
 (14, 0.12824729401064427),
 (15, 0.06282808624375433),
 (16, 0.07894736842105264),
 (17, 0.13977653617040256),
 (18, 0.09493290614465533),
```

In [ ]:
```

```

In [165]:
```
sorted(list(enumerate(similarity[0])), reverse=True, key=lambda x: x[1])[1:6]
```

Out[165]:
```
[(1216, 0.28676966733820225),
 (2409, 0.26901379342448517),
 (3731, 0.2605130246476754),
 (507, 0.255608593705383),
 (539, 0.25038669783359574)]
```

In [166]:
```
df.iloc[0]
```

Out[166]:
```
movie_id                                                19995
title                                                  Avatar
tags        in the 22nd century, a parapleg marin is dispa...
Name: 0, dtype: object
```

In [167]:
```
df.iloc[1216]
```

Out[167]:
```
movie_id                                                  440
title                                 Aliens vs Predator: Requiem
tags        a sequel to 2004' alien vs. predator, the icon...
Name: 1216, dtype: object
```

In [168]:
```
df.iloc[2409]
```

Out[168]:
```
movie_id                                                  679
title                                                 Aliens
tags        when ripley' lifepod is found by a salvag crew...
Name: 2409, dtype: object
```

In [ ]:
```

```

Final Function

In [202]:
```
def recommend(movie):

    movie_index = df[df['title'] == movie].index[0]
    distance = similarity[movie_index]
    movie_list = sorted(list(enumerate(distance)), reverse=True, key=lambda x: x[1])[1:6]

    for i in movie_list:
#         print(i[0])
        print(df.iloc[i[0]].title)
```

In [203]:
```
recommend('Avatar')
```

```
Aliens vs Predator: Requiem
Aliens
Falcon Rising
Independence Day
Titan A.E.
```

In [204]:
```
recommend('Iron Man')
```

```
Iron Man 3
Iron Man 2
Avengers: Age of Ultron
The Avengers
Captain America: Civil War
```

In [205]:
```
recommend('Captain America: Civil War')
```

```
Captain America: The First Avenger
Iron Man 3
Captain America: The Winter Soldier
Avengers: Age of Ultron
The Avengers
```

In [206]:
```
recommend('Spider-Man')
```

```
Spider-Man 3
Spider-Man 2
The Amazing Spider-Man 2
Arachnophobia
Kick-Ass
```

In [207]:
```
recommend('Superman')
```

```
Superman Returns
Superman II
Iron Man 2
Superman III
Superman IV: The Quest for Peace
```

In [208]:
```
recommend('X-Men')
```

```
X2
X-Men: The Last Stand
X-Men: Apocalypse
Iron Man 3
X-Men: First Class
```

In [ ]:
```

```

In [ ]:
```

```

In [171]:
```
similarity[0]
```

Out[171]:
```
array([1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
       0.        ])
```

In [172]:
```
df[df['title'] == 'Avatar']
```

Out[172]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...

In [174]:
```
df[df['title'] == 'Avatar'].index[0]
```

Out[174]:
```
0
```

In [196]:
```
df.iloc[1216]
```

Out[196]:
```
movie_id                                                  440
title                                 Aliens vs Predator: Requiem
tags        a sequel to 2004' alien vs. predator, the icon...
Name: 1216, dtype: object
```

In [197]:
```
df.iloc[1216].title
```

Out[197]: 'Aliens vs Predator: Requiem'

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: