In [1]:
```python
import numpy as np
import pandas as pd
```

In [2]:
```python
movies=pd.read_csv("tmdb_5000_movies.csv")
credits=pd.read_csv("tmdb_5000_credits.csv")
```

In [3]:
```python
movies=movies.merge(credits,on="title")
```

In [4]:
```python
movies.shape
```

Out[4]: (4809, 23)

In [5]:
```python
movies = movies[['movie_id','title','overview','genres','keywords','cast','
```

In [6]:
```python
movies.head()
```

Out[6]:

	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 "52fe48009251416c75(
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...	[{"cr "52fe4232c3a36847f8(
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...	[{"cr "54805967c3a36829b5(
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...	[{"cr "52fe4781c3a36847f81
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...	[{"cr "52fe479ac3a36847f81

```python
In [7]: movies.isnull().sum()
```

```
Out[7]: movie_id    0
        title       0
        overview    3
        genres      0
        keywords    0
        cast        0
        crew        0
        dtype: int64
```

```python
In [8]: #remove's all null value rows
        movies.dropna(inplace=True)
```

```python
In [9]: movies.shape
```

```
Out[9]: (4806, 7)
```

```python
In [10]: movies.duplicated().sum()
```

```
Out[10]: 0
```

```python
In [11]: movies.iloc[0].genres
```

```
Out[11]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 1
         4, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```python
In [12]: import ast
```

```python
In [13]: ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Advent
```

```
Out[13]: [{'id': 28, 'name': 'Action'},
          {'id': 12, 'name': 'Adventure'},
          {'id': 14, 'name': 'Fantasy'},
          {'id': 878, 'name': 'Science Fiction'}]
```

```python
In [14]: def convert(text):
             L = []
             for i in ast.literal_eval(text):
                 L.append(i['name'])
             return L
```

In [15]:
```python
movies['genres']=movies['genres'].apply(convert)
movies.head()
```

Out[15]:

	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", "...	[{"cre "52fe48009251416c750a
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...	[{"cre "52fe4232c3a36847f800
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...	[{"cre "54805967c3a36829b500
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...	[{"cre "52fe4781c3a36847f813
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...	[{"cre "52fe479ac3a36847f813

In [16]:
```python
movies['keywords']=movies['keywords'].apply(convert)
```

In [17]:
```python
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 [18]:
```python
movies['cast']=movies['cast'].apply(convert3)
```

In [19]: `movies.head()`

Out[19]:

	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 Weaver]	[{"cr" "52fe48009251416c750
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]	[{"cr" "52fe4232c3a36847f80
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]	[{"cr" "54805967c3a36829b50
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]	[{"cr" "52fe4781c3a36847f81
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]	[{"cr" "52fe479ac3a36847f81

In [20]:
```python
def fetch_director(text):
    L = []
    for i in ast.literal_eval(text):
        if i['job'] == 'Director':
            L.append(i['name'])
    return L
```

In [21]: `movies['crew']=movies['crew'].apply(fetch_director)`

In [22]: `movies.head()`

Out[22]:

	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 [23]:
```python
#to remove spaces
def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ",""))
    return L1
```

In [24]:
```python
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 [25]:
```python
movies.head()
```

Out[25]:

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesC
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...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreV
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[Sam
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[Christoph
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[Andrew

In [26]:
```python
movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

In [27]:
```python
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords']
```

In [28]:
```python
new_df = movies.drop(columns=['overview','genres','keywords','cast','crew']
```

In [29]:
```python
new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x))
new_df.head()
```

Out[29]:

	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 [30]:
```python
# ['Love','loving','lover']
# ['Love','Love','Love']
import nltk
```

In [31]:
```python
from nltk.stem.porter import PorterStemmer
ps=PorterStemmer()
```

In [32]:
```python
def stem(text):
    y=[]
    for i in text.split():
        y.append(ps.stem(i))
    return " ".join(y)
```

In [37]:
```python
new_df['tags']=new_df['tags'].apply(stem)
```

In [38]:
```python
new_df.head()
```

Out[38]:

	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 [41]:
```python
from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000,stop_words='english')
```

In [43]:
```python
vectors=cv.fit_transform(new_df['tags']).toarray()
```

In [46]:
```python
vectors
```

Out[46]:
```
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 [48]:
```python
cv.get_feature_names()
```

Out[48]:
```
['000',
 '007',
 '10',
 '100',
 '11',
 '12',
 '13',
 '14',
 '15',
 '16',
 '17',
 '17th',
 '18',
 '18th',
 '18thcenturi',
 '19',
 '1910',
 '1920',
 '1930',
```

In [49]:
```python
from sklearn.metrics.pairwise import cosine_similarity
```

In [50]:
```python
similarity = cosine_similarity(vectors)
```

In [51]:
```python
similarity
```

Out[51]:
```
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 [53]:
```python
sorted(list(enumerate(similarity[0])),reverse=True,key = lambda x: x[1])[1:
```

Out[53]:
```
[(1216, 0.2867696673382022),
 (2409, 0.26901379342448517),
 (3730, 0.2605130246476754),
 (507, 0.255608593705383),
 (539, 0.25038669783359574)]
```

```
In [77]: def recommend(movie):
             index = new_df[new_df['title'] == movie].index[0]
             distances = sorted(list(enumerate(similarity[index])),reverse=True,key
             for i in distances[1:6]:
                 print(new_df.iloc[i[0]].title)
```

```
In [79]: recommend('Batman Begins')
```

```
The Dark Knight
Batman
Batman
The Dark Knight Rises
10th & Wolf
```

```
In [ ]:
```