

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

In [3]:

```
movies.head(2)
```

Out[3]:

	budget	genres	homepage	id	keywords	origina
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	

In [4]:

```
movies.shape
```

Out[4]:

(4803, 20)

In [5]:

```
credits.head()
```

Out[5]:

	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 [6]:

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

In [7]:

```
movies.head()
```

Out[7]:

	budget	genres	homepage	id	keywords	orig
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1463, "name": "culture clash"}]	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name": "thriller"}]	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}]	http://www.thedarkknightises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "superhero"}]	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": 1463, "name": "culture clash"}]	

5 rows × 23 columns

In [8]:

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

In [9]:

```
movies.head()
```

Out[9]:

	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", "...
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...
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...
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...
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...

In [10]:

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

In [11]:

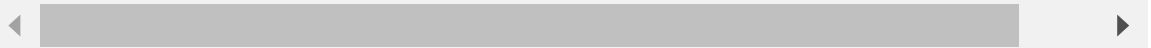
```
movies.dropna(inplace=True)
```

In [12]:

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

Out[12]:

	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...
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...
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...
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...
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...



In [13]:

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

Out[13]:

	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...	{{"cast_id": 242, "character": "Jake Sully", "...
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...
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...
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...
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...

In [14]:

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

Out[14]:

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

In [15]:

```
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 [16]:

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

Out[16]:

	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... {"ci": "52fe48009251416c75f8"}]
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, ... {"ci": "52fe4232c3a36847f8"}]
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... {"ci": "54805967c3a36829b5f8"}]
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... {"ci": "52fe4781c3a36847f8"}]
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,... {"ci": "52fe479ac3a36847f8"}]

In [17]:

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

In [18]:

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

In [19]:

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

In [20]:

```
movies.sample(5)
```

Out[20]:

	movie_id	title	overview	genres	keywords	cast	crew
688	693	Meet the Fockers	Hard-to-crack ex-CIA man, Jack Byrnes and his ...	[Comedy, Romance]	[cia, parents kids relationship, florida, anti...	[Ben Stiller, Teri Polo, Robert De Niro]	[Jay Roach]
47	54138	Star Trek Into Darkness	When the crew of the Enterprise is called back...	[Action, Adventure, Science Fiction]	[spacecraft, friendship, sequel, futuristic, s...	[Chris Pine, Zachary Quinto, Zoe Saldana]	[J.J. Abrams]
288	10545	The Hunchback of Notre Dame	When Quasi defies the evil Frollo and ventures...	[Drama, Animation, Family]	[paris, based on novel, judge, obsession, danc...	[Tom Hulce, Demi Moore, Tony Jay]	[Gary Trousdale, Kirk Wise]
1333	4944	Burn After Reading	When a disc containing memoirs of a former CIA...	[Comedy, Drama]	[blackmail, paranoia, fitness-training, plasti...	[George Clooney, Frances McDormand, Brad Pitt]	[Joel Coen, Ethan Coen]
2043	82650	Diary of a Wimpy Kid: Dog Days	School is out and Greg is ready for the days o...	[Comedy, Family]	[best friend, sweet sixteen, duringcreditsstin...	[Zachary Gordon, Robert Capron, Devon Bostick]	[David Bowers]

In [21]:

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

In [22]:

```
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 [23]:

```
movies.head()
```

Out[23]:

	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]
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]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]

In [24]:

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

In [25]:

```
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['ca
```

In [26]:

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

In [27]:

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

Out[27]:

	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 [28]:

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

In [29]:

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

In [30]:

```
vector.shape
```

Out[30]:

```
(4806, 5000)
```

In [31]:

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

In [32]:

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

In [33]:

```
similarity
```

Out[33]:

```
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 [34]:

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

Out[34]:

744

In [35]:

```
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)
```

In [36]:

```
recommend('Gandhi')
```

Gandhi, My Father
 The Wind That Shakes the Barley
 A Passage to India
 Guiana 1838
 Ramanujan

In [37]:

```
import pickle
```

In [38]:

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

In []:

In []: