

```
In [1]: import pandas as pd
import numpy as np
import warnings; warnings.simplefilter('ignore')

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

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

Out[3]:	budget	genres	homepage	id	keywords	original_language	origi
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...	en	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirat Ca At
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	Knig
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": ...	en	Jot

```
In [4]: movies.columns

Out[4]: 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 [5]: movies=movies[['genres', 'id', 'keywords', 'original_language',
'original_title', 'overview', 'popularity', 'production_companies', 'release_date', 'runt
'spoken_languages', 'tagline', 'title', 'vote_average',
'vote_count']]

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

Out[6]:	genres	id	keywords	original_language	original_title	overview	popularity	production_companies
0	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	19995	[[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]]	en	Avatar	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 those who have become his family.	150.437577	[[{"name": "Ingenious Film Partners", "id": 289}, {"name": "Lightstorm Entertainment", "id": 290}]]
1	[[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]]	285	[[{"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]]	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, has returned to bring the crew back to the Caribbean. The crew must work together to stop the evil forces that are threatening the world.	139.082615	[[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Piracy Pictures", "id": 3}]]
2	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	206647	[[{"id": 470, "name": "spy"}, {"id": 818, "name": "espionage"}]]	en	Spectre	A cryptic message from Bond's past sends him on a new mission.	107.376788	[[{"name": "Columbia Pictures", "id": 5}, {"name": "United Artists", "id": 6}]]
3	[[{"id": 28, "name": "Action"}, {"id": 80, "name": "Comic Book"}]]	49026	[[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "superhero"}]]	en	The Dark Knight Rises	Following the death of District Attorney Harvey Dent, Batman deduces that the only person who could have survived the attack is the Joker.	112.312950	[[{"name": "Legendary Pictures", "id": 923}, {"name": "Warner Bros. Entertainment", "id": 924}]]
4	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]]	49529	[[{"id": 818, "name": "based on novel"}, {"id": 819, "name": "science fiction"}]]	en	John Carter	John Carter is a war-weary, former military captain, plagued by traumatic experiences from a long and troubled past.	43.926995	[[{"name": "Walt Disney Pictures", "id": 2}, {"name": "United Artists", "id": 6}]]



```
In [7]: credits.head()
```

Out[7]:	movie_id	title	cast	crew
0	19995	Avatar	[[{"cast_id": 242, "character": "Jake Sully", "order": 1}, {"cast_id": 243, "character": "Neytiri", "order": 2}]]	[[{"credit_id": "52fe48009251416c750aca23", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 1}, {"credit_id": "52fe48009251416c750aca23", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 2}]]
1	285	Pirates of the Caribbean: At World's End	[[{"cast_id": 4, "character": "Captain Jack Sparrow", "order": 1}, {"cast_id": 5, "character": "Will Turner", "order": 2}]]	[[{"credit_id": "52fe4232c3a36847f800b579", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 1}, {"credit_id": "52fe4232c3a36847f800b579", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 2}]]
2	206647	Spectre	[[{"cast_id": 1, "character": "James Bond", "order": 1}, {"cast_id": 2, "character": "M", "order": 2}]]	[[{"credit_id": "54805967c3a36829b5002c41", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 1}, {"credit_id": "54805967c3a36829b5002c41", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 2}]]
3	49026	The Dark Knight Rises	[[{"cast_id": 2, "character": "Bruce Wayne / Batman", "order": 1}, {"cast_id": 3, "character": "Alfred Pennyworth", "order": 2}]]	[[{"credit_id": "52fe4781c3a36847f81398c3", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 1}, {"credit_id": "52fe4781c3a36847f81398c3", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 2}]]
4	49529	John Carter	[[{"cast_id": 5, "character": "John Carter", "order": 1}, {"cast_id": 6, "character": "Deeta", "order": 2}]]	[[{"credit_id": "52fe479ac3a36847f81398c3", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 1}, {"credit_id": "52fe479ac3a36847f81398c3", "department": "Production Design", "job": "Production Designer", "name": "Rick Heinrichs", "order": 2}]]

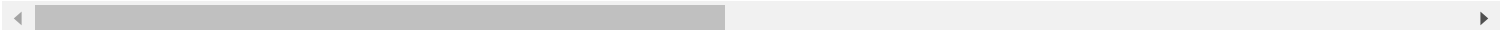
```
In [8]: credits.columns
```

```
Out[8]: Index(['movie_id', 'title', 'cast', 'crew'], dtype='object')
```

```
In [9]: credits=credits[['title', 'cast', 'crew']]
```

```
In [10]: df=pd.merge(movies,credits,on='title')
df.head()
```

Out[10]:	genres	id	keywords	original_language	original_title	overview	popularity	production_companies
0	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}], [{"id": 14, "name": "Action"}]]	19995	[[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}], [{"id": 1463, "name": "culture clash"}]]	en	Avatar	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 those who have become his family.	150.437577	[{"name": "Ingenious Film Partners", "id": 289}, {"name": "Lightstorm Entertainment", "id": 290}, {"name": "Twentieth Century Fox Film Corporation", "id": 291}]]
1	[[{"id": 12, "name": "Adventure"}], [{"id": 14, "name": "Action"}]]	285	[[{"id": 270, "name": "ocean"}, {"id": 726, "name": "nautilus"}], [{"id": 726, "name": "nautilus"}]]	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, has returned to bring the crew back to the Caribbean. The crew must work together to stop the evil forces that are trying to destroy the world.	139.082615	[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Buena Vista", "id": 3}, {"name": "Piracy", "id": 4}]]
2	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}]]	206647	[[{"id": 470, "name": "spy"}, {"id": 818, "name": "nameless"}], [{"id": 818, "name": "nameless"}]]	en	Spectre	A cryptic message from Bond's past sends him on a new mission.	107.376788	[{"name": "Columbia Pictures", "id": 5}, {"name": "United Artists", "id": 6}, {"name": "United Artists", "id": 7}]]
3	[[{"id": 28, "name": "Action"}, {"id": 80, "name": "nameless"}], [{"id": 80, "name": "nameless"}]]	49026	[[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "dc comics"}], [{"id": 853, "name": "dc comics"}]]	en	The Dark Knight Rises	Following the death of District Attorney Harvey Dent, Batman deduces that the only person who could have survived the attack is the Joker.	112.312950	[{"name": "Legendary Pictures", "id": 923}, {"name": "Warner Bros. Entertainment", "id": 924}, {"name": "Warner Bros. Entertainment", "id": 925}]]
4	[[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}], [{"id": 12, "name": "Adventure"}]]	49529	[[{"id": 818, "name": "based on novel"}, {"id": 819, "name": "based on novel"}], [{"id": 818, "name": "based on novel"}]]	en	John Carter	John Carter is a war-weary, former military captain, plagued by traumatic experiences from a long and troubled past.	43.926995	[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Buena Vista", "id": 3}, {"name": "Piracy", "id": 4}]]



```

In [11]: df.shape
Out[11]: (4809, 17)

In [12]: df.isna().sum()
Out[12]: genres      0
id      0
keywords    0
original_language  0
original_title  0
overview     3
popularity   0
production_companies  0
release_date  1
runtime      2
spoken_languages  0
tagline     844
title        0
vote_average  0
vote_count   0
cast         0
crew         0
dtype: int64

In [13]: df=df.dropna(subset='overview')
df.shape

```

Out[13]: (4806, 17)

```
In [14]: df.isna().sum()
```

Out[14]:

genres	0
id	0
keywords	0
original_language	0
original_title	0
overview	0
popularity	0
production_companies	0
release_date	1
runtime	0
spoken_languages	0
tagline	841
title	0
vote_average	0
vote_count	0
cast	0
crew	0

dtype: int64

```
In [15]: df.duplicated().sum()
```

Out[15]: 0

```
In [16]: df.head()
```

Out[16]:

	genres	id	keywords	original_language	original_title	overview	popularity	production_companies
0	[[{"id": 28, "name": "Action"}, {"id": 12, "nam...	19995	[[{"id": 1463, "name": "culture clash"}, {"id": ...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[[{"name": "Ingenious Film Partners", "id": 289...
1	[[{"id": 12, "name": "Adventure"}, {"id": 14, "...	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...	139.082615	[[{"name": "Walt Disney Pictures", "id": 2}, {""...
2	[[{"id": 28, "name": "Action"}, {"id": 12, "nam...	206647	[[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[[{"name": "Columbia Pictures", "id": 5}, {""nam...
3	[[{"id": 28, "name": "Action"}, {"id": 80, "nam...	49026	[[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[[{"name": "Legendary Pictures", "id": 923}, {""...
4	[[{"id": 28, "name": "Action"}, {"id": 12, "nam...	49529	[[{"id": 818, "name": "based on novel"}, {"id": ...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[[{"name": "Walt Disney Pictures", "id": 2}]]

```
In [40]: #data=df['id']
#data.head()
```

Out[40]:

0	19995
1	285
2	206647
3	49026
4	49529

Name: id, dtype: int64

```
In [41]: #data.to_csv('ids.csv',index=False)
```

```
In [18]: df.columns
```

Out[18]: Index(['genres', 'id', 'keywords', 'original_language', 'original_title', 'overview', 'popularity', 'production_companies', 'release_date', 'runtime', 'spoken_languages', 'tagline', 'title', 'vote_average', 'vote_count', 'cast', 'crew'], dtype='object')

```
In [19]: newDf=df[['genres', 'id', 'keywords',
'overview', 'title', 'cast', 'crew']]
```

```
In [20]: newDf.head()
```

Out[20]:

	genres	id	keywords	overview	title	cast	crew
0	["id": 28, "name": "Action"], {"id": 12, "nam...	19995	["id": 1463, "name": "culture clash"}, {"id":...	In the 22nd century, a paraplegic Marine is di...	Avatar	["cast_id": 242, "character": "Jake Sully", "...	["credit_id": "52fe48009251416c750aca23", "de...
1	["id": 12, "name": "Adventure"}, {"id": 14, "...	285	["id": 270, "name": "ocean"}, {"id": 726, "na...	Captain Barbossa, long believed to be dead, ha...	Pirates of the Caribbean: At World's End	["cast_id": 4, "character": "Captain Jack Spa...	["credit_id": "52fe4232c3a36847f800b579", "de...
2	["id": 28, "name": "Action"], {"id": 12, "nam...	206647	["id": 470, "name": "spy"}, {"id": 818, "name...	A cryptic message from Bond's past sends him o...	Spectre	["cast_id": 1, "character": "James Bond", "cr...	["credit_id": "54805967c3a36829b5002c41", "de...
3	["id": 28, "name": "Action"], {"id": 80, "nam...	49026	["id": 849, "name": "dc comics"}, {"id": 853,...	Following the death of District Attorney Harve...	The Dark Knight Rises	["cast_id": 2, "character": "Bruce Wayne / Ba...	["credit_id": "52fe4781c3a36847f81398c3", "de...
4	["id": 28, "name": "Action"], {"id": 12, "nam...	49529	["id": 818, "name": "based on novel"}, {"id":...	John Carter is a war-weary, former military ca...	John Carter	["cast_id": 5, "character": "John Carter", "c...	["credit_id": "52fe479ac3a36847f81398c3", "de...

```
In [21]: newDf['genres']=newDf['genres'].apply(lambda x: [i['name'] for i in eval(x)])
newDf['keywords']=newDf['keywords'].apply(lambda x: [i['name'] for i in eval(x)])
newDf['cast']=newDf['cast'].apply(lambda x: [i['name'] for i in eval(x)[:3]])
```

```
In [22]: def get_dir(crew_list):
for crew in eval(crew_list):
if crew['job'] == 'Director':
return crew['name']
return None
```

```
In [23]: newDf['crew'] = newDf['crew'].apply(get_dir)
```

```
In [24]: newDf.head()
```

Out[24]:

	genres	id	keywords	overview	title	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]	19995	[culture clash, future, space war, space colon...	In the 22nd century, a paraplegic Marine is di...	Avatar	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	James Cameron
1	[Adventure, Fantasy, Action]	285	[ocean, drug abuse, exotic island, east india ...	Captain Barbossa, long believed to be dead, ha...	Pirates of the Caribbean: At World's End	[Johnny Depp, Orlando Bloom, Keira Knightley]	Gore Verbinski
2	[Action, Adventure, Crime]	206647	[spy, based on novel, secret agent, sequel, mi...	A cryptic message from Bond's past sends him o...	Spectre	[Daniel Craig, Christoph Waltz, Léa Seydoux]	Sam Mendes
3	[Action, Crime, Drama, Thriller]	49026	[dc comics, crime fighter, terrorist, secret i...	Following the death of District Attorney Harve...	The Dark Knight Rises	[Christian Bale, Michael Caine, Gary Oldman]	Christopher Nolan
4	[Action, Adventure, Science Fiction]	49529	[based on novel, mars, medallion, space travel...	John Carter is a war-weary, former military ca...	John Carter	[Taylor Kitsch, Lynn Collins, Samantha Morton]	Andrew Stanton

```
In [25]: newDf['genres']=newDf['genres'].apply(lambda x : ' '.join([i.replace(" ",',') for i in x]))
newDf['keywords']=newDf['keywords'].apply(lambda x : ' '.join([i.replace(" ",',') for i in x]))
newDf['cast']=newDf['cast'].apply(lambda x : ' '.join([i.replace(" ",',') for i in x]))
newDf.head()
```

Out[25]:

	genres	id	keywords	overview	title	cast	crew
0	Action Adventure Fantasy ScienceFiction	19995	cultureclash future spacewar spacecolony socie...	In the 22nd century, a paraplegic Marine is di...	Avatar	SamWorthington ZoeSaldana SigourneyWeaver	James Cameron
1	Adventure Fantasy Action	285	ocean drugabuse exoticisland eastindiatradingc...	Captain Barbossa, long believed to be dead, ha...	Pirates of the Caribbean: At World's End	JohnnyDepp OrlandoBloom KeiraKnightley	Gore Verbinski
2	Action Adventure Crime	206647	spy basedonnovel secretagent sequel mi6 britis...	A cryptic message from Bond's past sends him o...	Spectre	DanielCraig ChristophWaltz LéaSeydoux	Sam Mendes
3	Action Crime Drama Thriller	49026	dccomics crimefighter terrorist secretidentity...	Following the death of District Attorney Harve...	The Dark Knight Rises	ChristianBale MichaelCaine GaryOldman	Christopher Nolan
4	Action Adventure ScienceFiction	49529	basedonnovel mars medallion spacetravel prince...	John Carter is a war-weary, former military ca...	John Carter	TaylorKitsch LynnCollins SamanthaMorton	Andrew Stanton

```
In [26]: newDf['tags']=newDf['genres'] + ' ' + newDf['keywords'] + ' ' + newDf['overview'] + ' ' + newDf[
newDf.head()
```

Out[26]:

	genres	id	keywords	overview	title	cast	crew	tags
0	Action Adventure Fantasy ScienceFiction	19995	cultureclash future spacewar spacecolony socie...	In the 22nd century, a paraplegic Marine is di...	Avatar	SamWorthington ZoeSaldana SigourneyWeaver	James Cameron	Action Adventure Fantasy ScienceFiction cultur...
1	Adventure Fantasy Action	285	ocean drugabuse exoticisland eastindiatradingc...	Captain Barbossa, long believed to be dead, ha...	Pirates of the Caribbean: At World's End	JohnnyDepp OrlandoBloom KeiraKnightley	Gore Verbinski	Adventure Fantasy Action ocean drugabuse exoti...
2	Action Adventure Crime	206647	spy basedonnovel secretagent sequel mi6 britis...	A cryptic message from Bond's past sends him o...	Spectre	DanielCraig ChristophWaltz LéaSeydoux	Sam Mendes	Action Adventure Crime spy basedonnovel secret...
3	Action Crime Drama Thriller	49026	dccomics crimefighter terrorist secretidentity...	Following the death of District Attorney Harve...	The Dark Knight Rises	ChristianBale MichaelCaine GaryOldman	Christopher Nolan	Action Crime Drama Thriller dccomics crimefigh...
4	Action Adventure ScienceFiction	49529	basedonnovel mars medallion spacetravel prince...	John Carter is a war- weary, former military ca...	John Carter	TaylorKitsch LynnCollins SamanthaMorton	Andrew Stanton	Action Adventure ScienceFiction basedonnovel m...

In [27]:

```
finalDf=newDf[['id','title','tags']]  
finalDf.head()
```

Out[27]:

	id	title	tags
0	19995	Avatar	Action Adventure Fantasy ScienceFiction cultur...
1	285	Pirates of the Caribbean: At World's End	Adventure Fantasy Action ocean drugabuse exoti...
2	206647	Spectre	Action Adventure Crime spy basedonnovel secret...
3	49026	The Dark Knight Rises	Action Crime Drama Thriller dccomics crimefigh...
4	49529	John Carter	Action Adventure ScienceFiction basedonnovel m...

In [28]:

```
import nltk as nl  
from nltk.stem.porter import PorterStemmer  
ps = PorterStemmer()
```

In [29]:

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

In [30]:

```
finalDf['tags'] = finalDf['tags'].astype(str)
```

In [31]:

```
finalDf['tags']=finalDf['tags'].apply(stem)
```

In [32]:

```
finalDf.head()
```

		id	title	tags
0	19995		Avatar	action adventur fantasi sciencefict culturecla...
1	285	Pirates of the Caribbean: At World's End		adventur fantasi action ocean drugabus exotici...
2	206647		Spectre	action adventur crime spi basedonnovel secreta...
3	49026	The Dark Knight Rises		action crime drama thriller dccomic crimefight...
4	49529		John Carter	action adventur sciencefict basedonnovel mar m...

```
In [33]: from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000, stop_words='english')
vectors=cv.fit_transform(finalDf['tags']).toarray()
```

```
In [ ]:
```

```
In [34]: vectors[0]
```

```
Out[34]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
In [35]: from sklearn.metrics.pairwise import cosine_similarity
similarity = cosine_similarity(vectors)
```

```
In [36]: def recommend(movie):
movie_index=finalDf[finalDf['title']==movie].index[0]
distance=similarity[movie_index]
movie_list = sorted(list(enumerate(distance)), reverse=True, key=lambda x:x[1])[1:10]

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

```
In [37]: recommend('Batman Begins')
```

```
The Dark Knight
Batman
Batman
The Dark Knight Rises
Batman Returns
Batman v Superman: Dawn of Justice
Batman & Robin
10th & Wolf
Rockaway
```

```
In [38]: recommend('Iron Man')
```

```
Iron Man 2
Iron Man 3
Avengers: Age of Ultron
Captain America: Civil War
The Avengers
Ant-Man
Guardians of the Galaxy
X-Men
X-Men Origins: Wolverine
```