

```
In [3]: import numpy as np
import pandas as pd
```

```
In [4]: movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

```
In [5]: movies.head()
```

```
Out[5]:
```

	budget	genres	homepage	id	keywords	original_title
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1463, "name": "culture clash"}]	Avatar
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"}]	Pirates of the Caribbean: The Curse of the Black Pearl
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name": "thriller"}]	Spectre
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}, {"id": 96, "name": "Drama"}]	http://www.thedarkknighttrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "superhero"}]	The Dark Knight Rises
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": 1463, "name": "culture clash"}]	John Carter

```
In [6]: movies = movies.merge(credits,on='title')
```

In [7]:

movies.head(1)

Out[7]:

	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 rows × 23 columns

In [8]:

```
# COLUMNS TO KEEP
# genres
# id
# keywords
# title
# overview
# cast
# crew
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]: movies.isnull().sum()
```

```
Out[10]: movie_id    0
         title       0
         overview    3
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [11]: movies.dropna(inplace=True)
```

```
In [12]: movies.isnull().sum()
```

```
Out[12]: movie_id    0
         title       0
         overview    0
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [13]: movies.duplicated().sum()
```

```
Out[13]: 0
```

```
In [14]: movies.iloc[0].genres
```

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

```
In [15]: import ast
         def convert(obj):
             L=[]
             for i in ast.literal_eval(obj):
                 L.append(i['name'])
             return L
```

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

```
In [17]: movies['keywords']=movies['keywords'].apply(convert)
```

```
In [18]: def convert3(obj):
          L=[]
          counter=0
          for i in ast.literal_eval(obj):
              if counter != 3:
                  L.append(i['name'])
              else:
                  break
          return L
```

```
In [19]: movies['cast']=movies['cast'].apply(convert3)
```

```
In [20]: movies.head(1)
```

```
Out[20]:
```

	movie_id	title	overview	genres	keywords	cast	cre
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...	["credit_id": "52fe48009251416c750aca23", "de

```
In [21]: def fetch_director(obj):
          L=[]
          for i in ast.literal_eval(obj):
              if i['job']=='Director':
                  L.append(i['name'])
              break
          return L
```

```
In [22]: movies['crew']=movies['crew'].apply(fetch_director)
```

```
In [23]: movies['overview']=movies['overview'].apply(lambda x: x.split())
```

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

Out[24]:

	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 Weave...	[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, R...	[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, A...	[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 [25]: `movies['genres']=movies['genres'].apply(lambda x: [i.replace(" ",",") for i in x])`  
`movies['keywords']=movies['keywords'].apply(lambda x: [i.replace(" ",",") for i in x])`  
`movies['cast']=movies['cast'].apply(lambda x: [i.replace(" ",",") for i in x])`  
`movies['crew']=movies['crew'].apply(lambda x: [i.replace(" ",",") for i in x])`

In [26]:

movies.head()

Out[26]:

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[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, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste...
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux, Ralp...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman, Anne...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton, Wi...

In [27]:

movies['tags'] = movies['overview'] + movies['genres']+movies['keywords']+m

```
In [28]: movies.head()
```

Out[28]:

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[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, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste...
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux, Ralp...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman, Anne...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton, Wi...

```
In [29]: new_df=movies[['movie_id','title','tags']]
```

```
In [30]: new_df
```

Out[30]:

	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...

4806 rows × 3 columns

```
In [31]: new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

C:\Users\Dell\AppData\Local\Temp\ipykernel\_16772\487797088.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

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

```
In [32]: new_df.head()
```

```
Out[32]:
```

	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 [33]: new_df['tags']=new_df['tags'].apply(lambda x:x.lower())
```

C:\Users\Dell\AppData\Local\Temp\ipykernel\_16772\4224080999.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

```
new_df['tags']=new_df['tags'].apply(lambda x:x.lower())
```

```
In [35]: import nltk
```

```
In [36]: from nltk.stem.porter import PorterStemmer
ps=PorterStemmer()
```

```
In [37]: def stem(text):
y=[]
for i in text.split():
y.append(ps.stem(i))
return " ".join(y)
```



```
In [38]: new_df['tags']=new_df['tags'].apply(stem)
```

C:\Users\Dell\AppData\Local\Temp\ipykernel\_16772\3514595201.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

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

## Vectorization

```
In [39]: from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000, stop_words='english')
```

```
In [40]: vectors=cv.fit_transform(new_df['tags']).toarray()
```

```
In [41]: vectors[0]
```

```
Out[41]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
In [42]: cv.get_feature_names_out()
```

```
Out[42]: array(['000', '007', '10', ..., 'zoo', 'zooeydeschanel', 'zoëkravitz'],
dtype=object)
```

```
In [43]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [44]: similarity=cosine_similarity(vectors)
```

```
In [45]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]
```

```
Out[45]: [(1216, 0.27028123880866767),
(582, 0.23162743094465488),
(2333, 0.22996655275195),
(3730, 0.22498852128662872),
(507, 0.22438727760202976)]
```

```
In [71]: def recommend(movie):
    movie_index=new_df[new_df['title']==movie].index[0]
    distances=similarity[movie_index]
    movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x

    for i in movies_list:
        print(new_df.iloc[i[0]].title)
```

```
In [72]: recommend('The Dark Knight Rises')
```

```
The Dark Knight
Batman Begins
Batman Returns
Batman
Batman Forever
```

```
In [56]: import pickle
```

```
In [57]: pickle.dump(new_df, open('movies.pkl', 'wb'))
```

```
In [58]: pickle.dump(similarity, open('similarity.pkl', 'wb'))
```

```
In [59]: pickle.dump(new_df.to_dict(), open('movie_dict.pkl', 'wb'))
```

```
In [62]: mov = pd.DataFrame(new_df.to_dict())
```

```
In [63]: mov
```

```
Out[63]:
```

	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...

```
4806 rows × 3 columns
```

```
In [ ]:
```