

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

```
In [6]: movies=pd.read_csv('C:\\Users\\waghm\\OneDrive\\Desktop\\Movies\\tmdb_5000_
```

```
In [7]: movies.head(5)
```

```
Out[7]:
```

	budget	genres	homepage	id	keywords	origi
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...	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknighttrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	

In [8]: `movies.describe()`

Out[8]:

	budget	id	popularity	revenue	runtime	vote_average
<b>count</b>	4.803000e+03	4803.000000	4803.000000	4.803000e+03	4801.000000	4803.000000
<b>mean</b>	2.904504e+07	57165.484281	21.492301	8.226064e+07	106.875859	6.092172
<b>std</b>	4.072239e+07	88694.614033	31.816650	1.628571e+08	22.611935	1.194612
<b>min</b>	0.000000e+00	5.000000	0.000000	0.000000e+00	0.000000	0.000000
<b>25%</b>	7.900000e+05	9014.500000	4.668070	0.000000e+00	94.000000	5.600000
<b>50%</b>	1.500000e+07	14629.000000	12.921594	1.917000e+07	103.000000	6.200000
<b>75%</b>	4.000000e+07	58610.500000	28.313505	9.291719e+07	118.000000	6.800000
<b>max</b>	3.800000e+08	459488.000000	875.581305	2.787965e+09	338.000000	10.000000

In [9]: `movies.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   budget                4803 non-null   int64
1   genres                4803 non-null   object
2   homepage              1712 non-null   object
3   id                    4803 non-null   int64
4   keywords              4803 non-null   object
5   original_language     4803 non-null   object
6   original_title        4803 non-null   object
7   overview              4800 non-null   object
8   popularity            4803 non-null   float64
9   production_companies  4803 non-null   object
10  production_countries  4803 non-null   object
11  release_date          4802 non-null   object
12  revenue               4803 non-null   int64
13  runtime               4801 non-null   float64
14  spoken_languages      4803 non-null   object
15  status                4803 non-null   object
16  tagline               3959 non-null   object
17  title                 4803 non-null   object
18  vote_average          4803 non-null   float64
19  vote_count            4803 non-null   int64
dtypes: float64(3), int64(4), object(13)
memory usage: 750.6+ KB
```

```
In [11]: movies.isnull().sum()
```

```
Out[11]: budget          0
genres              0
homepage          3091
id                 0
keywords           0
original_language  0
original_title     0
overview           3
popularity         0
production_companies 0
production_countries 0
release_date       1
revenue            0
runtime            2
spoken_languages   0
status             0
tagline            844
title              0
vote_average       0
vote_count         0
dtype: int64
```

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

```
Out[12]: 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 [15]: movies=movies[['id','title','genres','overview']]
```

In [16]: movies

Out[16]:

	id	title	genres	overview
0	19995	Avatar	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	A cryptic message from Bond's past sends him o...
3	49026	The Dark Knight Rises	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	Following the death of District Attorney Harve...
4	49529	John Carter	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	John Carter is a war-weary, former military ca...
...	...	...	...	...
4798	9367	El Mariachi	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	El Mariachi just wants to play his guitar and ...
4799	72766	Newlyweds	[{"id": 35, "name": "Comedy"}, {"id": 10749, "...	A newlywed couple's honeymoon is upended by th...
4800	231617	Signed, Sealed, Delivered	[{"id": 35, "name": "Comedy"}, {"id": 18, "nam...	"Signed, Sealed, Delivered" introduces a dedic...
4801	126186	Shanghai Calling		When ambitious New York attorney Sam is sent t...
4802	25975	My Date with Drew	[{"id": 99, "name": "Documentary"}]	Ever since the second grade when he first saw ...

4803 rows × 4 columns

In [18]: movies['tags']=movies['overview']+movies['genres']

C:\Users\waghm\AppData\Local\Temp\ipykernel\_15620\2350984780.py:1: Setting WithCopyWarning:  
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)

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

In [19]: movies

Out[19]:

	id	title	genres	overview	tags
0	19995	Avatar	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	In the 22nd century, a paraplegic Marine is di...	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}]	Captain Barbossa, long believed to be dead, ha...	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	A cryptic message from Bond's past sends him o...	A cryptic message from Bond's past sends him o...
3	49026	The Dark Knight Rises	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Thriller"}]	Following the death of District Attorney Harve...	Following the death of District Attorney Harve...
4	49529	John Carter	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	John Carter is a war-weary, former military ca...	John Carter is a war-weary, former military ca...
...	...	...	...	...	...
4798	9367	El Mariachi	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Thriller"}]	El Mariachi just wants to play his guitar and ...	El Mariachi just wants to play his guitar and ...
4799	72766	Newlyweds	[{"id": 35, "name": "Comedy"}, {"id": 10749, "name": "Romance"}]	A newlywed couple's honeymoon is upended by th...	A newlywed couple's honeymoon is upended by th...
4800	231617	Signed, Sealed, Delivered	[{"id": 35, "name": "Comedy"}, {"id": 18, "name": "Action"}]	"Signed, Sealed, Delivered" introduces a dedic...	"Signed, Sealed, Delivered" introduces a dedic...
4801	126186	Shanghai Calling	[]	When ambitious New York attorney Sam is sent t...	When ambitious New York attorney Sam is sent t...
4802	25975	My Date with Drew	[{"id": 99, "name": "Documentary"}]	Ever since the second grade when he first saw ...	Ever since the second grade when he first saw ...

4803 rows × 5 columns

In [22]: new\_data=movies.drop(columns=['overview', 'genres'])

```
In [23]: new_data
```

```
Out[23]:
```

	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...
...	...		...	...
4798	9367		El Mariachi	El Mariachi just wants to play his guitar and ...
4799	72766		Newlyweds	A newlywed couple's honeymoon is upended by th...
4800	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...	
4801	126186		Shanghai Calling	When ambitious New York attorney Sam is sent t...
4802	25975		My Date with Drew	Ever since the second grade when he first saw ...

4803 rows × 3 columns

```
In [24]: from sklearn.feature_extraction.text import CountVectorizer
```

```
In [26]: cv=CountVectorizer(max_features=1000,stop_words='english')
```

```
In [27]: cv
```

```
Out[27]: CountVectorizer(max_features=1000, stop_words='english')
```

```
In [30]: vector=cv.fit_transform(new_data['tags'].values.astype('U')).toarray()
```

```
In [31]: vector.shape
```

```
Out[31]: (4803, 1000)
```

```
In [32]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [33]: similarity=cosine_similarity(vector)
```

```
In [34]: similarity
```

```
Out[34]: array([[1.          , 0.70064905, 0.52466067, ..., 0.43070552, 0.
        0.17712298],
        [0.70064905, 1.          , 0.49779573, ..., 0.4086515 , 0.
        0.15512631],
        [0.52466067, 0.49779573, 1.          , ..., 0.34425764, 0.
        0.13068205],
        ...,
        [0.43070552, 0.4086515 , 0.34425764, ..., 1.          , 0.09028939,
        0.17879963],
        [0.          , 0.          , 0.          , ..., 0.09028939, 1.          ,
        0.09901475],
        [0.17712298, 0.15512631, 0.13068205, ..., 0.17879963, 0.09901475,
        1.          ]])
```

```
In [36]: new_data[new_data['title']=="The Godfather"].index[0]
```

```
Out[36]: 3337
```

```
In [40]: distance = sorted(list(enumerate(similarity[3337])),reverse=True,key=lambda
for i in distance[0:5]:
    print(new_data.iloc[i[0]].title)
```

```
The Godfather
The Godfather: Part II
Nine Queens
Mi America
Vera Drake
```

```
In [44]: def recommnad(movies):
        index=new_data[new_data['title']==movies].index[0]
        distance = sorted(list(enumerate(similarity[index])),reverse=True,key=l
        for i in distance[0:5]:
            print(new_data.iloc[i[0]].title)
```

```
In [45]: recommnad("Iron Man")
```

```
Iron Man
Godzilla 2000
Mad Max Beyond Thunderdome
Star Wars: Episode I - The Phantom Menace
X-Men: The Last Stand
```

```
In [47]: import pickle
```

```
In [48]: pickle.dump(new_data,open('movies_list.pkl','wb'))
```

```
In [50]: pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [51]: pickle.load(open('movies_list.pkl', 'rb'))
```

Out[51]:

	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...
...	...		...	...
4798	9367		El Mariachi	El Mariachi just wants to play his guitar and ...
4799	72766		Newlyweds	A newlywed couple's honeymoon is upended by th...
4800	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...	
4801	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...	
4802	25975	My Date with Drew	Ever since the second grade when he first saw ...	

4803 rows × 3 columns

```
In [ ]:
```