

```
[1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings as w
warnings.filterwarnings("ignore")

# For mathematical calculations
# For data visualization
# For plotting graphs
# To ignore any warnings

In [2]: filepath = "C:\Users\91623\OneDrive\Desktop\projects\movie\movie.csv"
df = pd.read_csv(filepath)
print(df)
```

```
   Movie_ID  Movie_Title  Movie_Genre \
0         1      Four Rooms      Crime Comedy
1         2      Star Wars      Adventure Action Science Fiction
2         3      Finding Memo      Animation Family
3         4      Forrest Gump      Comedy Drama Romance
4         5      American Beauty      Drama
...      ...      ...      ...
4755      4755      Midnight Cabaret      Horror
4756      4756      Growing Up Smith      Comedy Family Drama
4757      4757      8 Days      Thriller Drama
4758      4758      Running Forever      Family
4759      4758      To Be Frank, Sinatra at 100      Documentary

   Movie_Language  Movie_Budget  Movie_Popularity  Movie_Release_Date \
0                en      4000000      22.876230      09-12-1995
1                en      1000000      126.393695      25-03-1977
2                en      94000000      85.688789      30-05-2003
3                en      55000000      138.133331      06-07-1994
4                en      15000000      80.878605      15-09-1999
...      ...      ...      ...
4755              en      0      0.002195      01-01-1990
4756              en      0      0.719870      03-02-2017
4757              en      0      0.010295      15-06-2014
4758              en      0      0.020756      27-10-2013
4759              en      2      0.050625      12-12-2013

   Movie_Revenue  Movie_Runtime  Movie_Vote  ... \
0      43000000      98.0      6.5      ...
1      773398007      121.0      8.7      ...
2      940335336      100.0      7.6      ...
3      677945399      142.0      8.2      ...
4      356296601      122.0      6.9      ...
...      ...      ...      ...
4755              0      94.0      0.0      ...
4756              0      102.0      7.4      ...
4757              0      90.0      0.0      ...
4758              0      88.0      0.0      ...
4759              0      NaN      0.0      ...

   Movie_Homepage \
0                NaN
1      http://www.starwars.com/films/star-wars-episode...
2      http://movies.disney.com/finding-nemo
3                NaN
4      http://www.dreamworks.com/ab/
...
4755                NaN
4756      http://www.growingupsmithmovie.com
4757                NaN
4758                NaN
4759                NaN

   Movie_Keywords \
0      hotel new year's eve witch bet hotel room
1      android galaxy hermit death star lightaber
2      father son relationship harbor underwater fish...
3      comedy veteran hippie mentally disabled runni...
4      male nudity female nudity adultery midlife cri...
...
4755                NaN
4756                NaN
4757      christian film sex trafficking
4758                NaN
4759      music actors legendary performer classic hollyw...

   Movie_Overview \
0      It's Ted the Bellhop's first night on the job...
1      Princess Leia is captured and held hostage by ...
2      Memo, an adventurous young clownfish, is unexp...
3      A man with a low IQ has accomplished great thi...
4      Lester Burnham, a depressed suburban father in...
...
4755      A Broadway producer puts on a play with a Devil...
4756      In 1979, an Indian family moves to America wit...
4757      After sneaking to a party with her friends, 16...
4758      After being estranged since her mother's death...
4759                NaN

   Movie_Production_House \
0      [{"name": "Miramax Films", "id": 14}, {"name": "...
1      [{"name": "Lucasfilm", "id": 1}, {"name": "The...
2      [{"name": "Pixar Animation Studios", "id": 3}]
3      [{"name": "Paramount Pictures", "id": 4}]
4      [{"name": "DreamWorks SKG", "id": 27}, {"name": "...
...
4755                ...
4756                ...
4757      [{"name": "After Eden Pictures", "id": 85248}]
4758      [{"name": "New Kingdom Pictures", "id": 41673}]
4759      [{"name": "Eyeline Entertainment", "id": 60343}]

   Movie_Production_Country \
0      [{"iso_3166_1": "US", "name": "United States o...
1      [{"iso_3166_1": "US", "name": "United States o...
2      [{"iso_3166_1": "US", "name": "United States o...
3      [{"iso_3166_1": "US", "name": "United States o...
4      [{"iso_3166_1": "US", "name": "United States o...
...
4755                ...
4756                ...
4757      [{"iso_3166_1": "US", "name": "United States o...
4758      [{"iso_3166_1": "US", "name": "United States o...
4759      [{"iso_3166_1": "Gb", "name": "United Kingdom"}]

   Movie_Spoken_Language \
0      [{"iso_639_1": "en", "name": "English"}]
1      [{"iso_639_1": "en", "name": "English"}]
2      [{"iso_639_1": "en", "name": "English"}]
3      [{"iso_639_1": "en", "name": "English"}]
4      [{"iso_639_1": "en", "name": "English"}]
...
4755      [{"iso_639_1": "en", "name": "English"}]
4756      [{"iso_639_1": "en", "name": "English"}]
4757      [{"iso_639_1": "en", "name": "English"}]
4758                []
4759                []

   Movie_Tagline \
0      Twelve outrageous guests. Four scandalous requ...
1      A long time ago in a galaxy far, far away...
2      There are 3.7 trillion fish in the ocean, they...
3      The world will never be the same, once you've ...
4      Look closer.
...
4755                ...
4756      It's better to stand out than to fit in.
4757      She never knew it could happen to her...
4758                NaN
4759                NaN

   Movie_Cast \
0      Tim Roth Antonio Banderas Jennifer Beals Madon...
1      Mark Hamill Harrison Ford Carrie Fisher Peter ...
2      Albert Brooks Ellen DeGeneres Alexander Gould...
3      Tom Hanks Robin Wright Gary Sinise Mykelti Will...
...
4755      Lisa Hart Carroll Michael Des Barres Paul Drak...
4756      Roni Akurani Brighton Sharbino Jason Lee Anj...
4757      Nicole Smolen Kim Baldwin Ariana Stephens Brys...
4758                NaN
4759      Tony Oppedisano

   Movie_Crew \
0      [{"name": "Allison Anders", "gender": 1, "depa...
1      [{"name": "George Lucas", "gender": 2, "depart...
2      [{"name": "Andrew Stanton", "gender": 3, "depa...
3      [{"name": "Alan Silvestri", "gender": 2, "depa...
...
4755      [{"name": "Peece Dingoo", "gender": 0, "departme...
4756      [{"name": "Frank Lotito", "gender": 0, "depart...
4757      [{"name": "Jaco Booyens", "gender": 0, "depart...
4758      [{"name": "Simon Napier-Bell", "gender": 0, "d...
4759                NaN

[4760 rows x 21 columns]
```

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

```
Out[3]:
```

	Movie_ID	Movie_Title	Movie_Genre	Movie_Language	Movie_Budget	Movie_Popularity	Movie_Release_Date	Movie_Revenue	Movie_Runtime
0	1	Four Rooms	Crime Comedy	en	4000000	22.876230	09-12-1995	4300000	98
1	2	Star Wars	Adventure Action Science Fiction	en	11000000	126.393695	25-03-1977	775398007	121
2	3	Finding Memo	Animation Family	en	94000000	85.688789	30-05-2003	940335336	100
3	4	Forrest Gump	Comedy Drama Romance	en	55000000	138.133331	06-07-1994	677945399	142
4	5	American Beauty	Drama	en	15000000	80.878605	15-09-1999	356296601	122

5 rows x 21 columns

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4760 entries, 0 to 4759
Data columns (total 21 columns):
 #   Column              Non-Null Count  Dtype
---  --
 0   Movie_ID            4760 non-null   int64
 1   Movie_Title         4760 non-null   object
 2   Movie_Genre         4760 non-null   object
 3   Movie_Language      4760 non-null   object
 4   Movie_Budget        4760 non-null   int64
 5   Movie_Popularity     4760 non-null   float64
 6   Movie_Release_Date  4760 non-null   object
 7   Movie_Revenue       4760 non-null   int64
 8   Movie_Runtime       4760 non-null   float64
 9   Movie_Vote          4760 non-null   float64
10   Movie_Vote_Count    4760 non-null   int64
11   Movie_Homepage      4760 non-null   object
12   Movie_Keywords      4373 non-null   object
13   Movie_Overview      4697 non-null   object
14   Movie_Production_House 4760 non-null   object
15   Movie_Production_Country 4760 non-null   object
16   Movie_Spoken_Language 4760 non-null   object
17   Movie_Tagline       3942 non-null   object
18   Movie_Cast          4753 non-null   object
19   Movie_Crew          4760 non-null   object
20   Movie_Director       4738 non-null   object
dtype: object
memory usage: 781.1+ KB
```

```
In [5]: df.shape
```

```
Out[5]: (4760, 21)
```

```
In [6]: feat = df[['Movie_Genre', 'Movie_Keywords', 'Movie_Tagline', 'Movie_Cast', 'Movie_Director']].fillna('')
```

```
In [7]: feat
```

```
Out[7]:
```

	Movie_Genre	Movie_Keywords	Movie_Tagline	Movie_Cast	Movie_Director
0	Crime Comedy	hotel new year's eve witch bet hotel room	Twelve outrageous guests. Four scandalous requ...	Tim Roth Antonio Banderas Jennifer Beals Madon...	Allison Anders
1	Adventure Action Science Fiction	android galaxy hermit death star lightaber	A long time ago in a galaxy far, far away...	Mark Hamill Harrison Ford Carrie Fisher Peter ...	George Lucas
2	Animation Family	father son relationship harbor underwater fish...	There are 3.7 trillion fish in the ocean, they...	Albert Brooks Ellen DeGeneres Alexander Gould...	Andrew Stanton
3	Comedy Drama Romance	comedy veteran hippie mentally disabled runni...	The world will never be the same, once you've ...	Tom Hanks Robin Wright Gary Sinise Mykelti Will...	Robert Zemeckis
4	Drama	male nudity female nudity adultery midlife cri...	Look closer.	Kevin Spacey Annette Bening Thora Birch Wes Be...	Sam Mendes
...
4755	Horror	The hot spot where Satan's waitin'.	The hot spot where Satan's waitin'.	Lisa Hart Carroll Michael Des Barres Paul Drak...	Peece Dingoo
4756	Comedy Family Drama	It's better to stand out than to fit in.	It's better to stand out than to fit in.	Roni Akurani Brighton Sharbino Jason Lee Anj...	Frank Lotito
4757	Thriller Drama	christian film sex trafficking	She never knew it could happen to her...	Nicole Smolen Kim Baldwin Ariana Stephens Brys...	Jaco Booyens
4758	Family				
4759	Documentary	music actors legendary performer classic hollyw...		Tony Oppedisano	Simon Napier-Bell

4760 rows x 5 columns

```
In [8]: x = feat[['Movie_Genre'] + ' ' + feat['Movie_Keywords'] + ' ' + feat['Movie_Tagline'] + ' ' + feat['Movie_Cast'] + ' ' + feat['Movie_Director']]
```

```
In [9]: x
```

```
Out[9]:
```

```
0      Crime Comedy hotel new year's eve witch bet ho...
1      Adventure Action Science Fiction android galax...
2      Animation Family father son relationship harb...
3      Comedy Drama Romance comedy veteran hippie me...
4      Drama male nudity female nudity adultery midli...
...
4755      Horror The hot spot where Satan's waitin'. L...
4756      Comedy Family Drama It's better to stand ou...
4757      Thriller Drama christian film sex trafficking...
4758      Family
4759      Documentary music actors legendary performer c...
Length: 4760, dtype: object
```

```
In [10]:
```

```
In [10]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [11]: tfidf = TfidfVectorizer()
```

```
In [12]: x = tfidf.fit_transform(x)
```

```
In [13]: print(x)
```

```
(0, 617)      0.1633382144607913
(0, 692)      0.1632591540386865
(0, 15413)     0.1465525095337543
(0, 9675)      0.1422057295252661
(0, 8465)      0.1653984137820977
(0, 1390)      0.16898383612799558
(0, 8625)      0.03979561597509463
(0, 1214)      0.13865957545146072
(0, 729)       0.1341506339531618
(0, 13053)     0.1432591540386865
(0, 15355)     0.104778192972666779
(0, 9048)      0.0868642116160778
(0, 11611)     0.06265380151644369
(0, 16773)     0.17654247479915475
(0, 8625)      0.03979561597509463
(0, 16735)     0.106900837515254919
(0, 7904)      0.13348000542112332
(0, 15219)     0.0800472886454354
(0, 11242)     0.072777982394848746
(0, 9878)      0.11398399582562203
(0, 14899)     0.1145405151030811
(0, 7071)      0.198224175984064614
(0, 7454)      0.14746357854121262
(0, 1495)      0.197126373873161423
(0, 9206)      0.15186283580984414
...
(4757, 5455)  0.12491480594769522
(4757, 2367)  0.162737583631626
(4757, 4464)  0.2352246355046633
(4757, 6938)  0.17088136781316628
(4757, 6379)  0.17486028595721913
(4757, 15303) 0.07654356007668191
(4757, 15384) 0.09754322497537371
(4757, 7648)  0.11474621494364192
(4757, 10896) 0.14546473055064447
(4757, 4494)  0.05675298448720501
(4758, 5238)  1.0
(4759, 11264) 0.33947721804318337
(4759, 11708) 0.33947721804318337
(4759, 205)   0.32379181628497312
(4759, 8902)  0.3040280704566037
(4759, 14062) 0.32379181628497312
(4759, 3058)  0.2812896191863103
(4759, 7130)  0.26416624494963793
(4759, 10761) 0.312617595732147
(4759, 4358)  0.18306542312175342
(4759, 14051) 0.2008431377466435
(4759, 5690)  0.19534291014627303
(4759, 15431) 0.19628653185946862
(4759, 1490)  0.21197254705920862
(4759, 10666) 0.1588826898734043
```

```
In [15]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [16]: sim_scr = cosine_similarity(x)
```

```
In [17]: sim_scr
```

```
Out[17]: array([[1.         , 0.0351235, 0.03570468, ..., 0.         , 0.         ,
0.0351235, 1.         , 0.00806674, ..., 0.         , 0.         ,
0.         , 0.03570468, 0.00806674, ..., 0.         , 0.08014876,
0.         , 1.         ],
[0.03570468, 0.         , 0.0806674, 1.         , ..., 0.         , 0.08014876,
0.         , 1.         ],
...,
[0.         , 0.         , 0.         , ..., 1.         , 0.         ,
0.         , 1.         , 0.08014876, ..., 0.         , 1.         ,
0.         , 1.         ],
[0.         , 0.         , 0.         , ..., 0.         , 0.         ,
1.         , 1.         ]])
```

```
In [18]: fav_movie = input('enter')
enteravtar
```

```
In [19]: all = df[['Movie_Title']].tolist()
```

```
In [21]: import difflib
```

```
In [23]: rec = difflib.get_close_matches(fav_movie,all)
print(rec)
```

```
['Avatar', 'Salvador', 'Water']
```

```
In [24]: close_match = rec[0]
print(close_match)
Avatar
```

```
In [30]: index_close = df[df.Movie_Title == close_match]['Movie_ID'].values[0]
print(index_close)
2692
```

getting list of similar movie

```
In [33]: rec_score = list(enumerate(sim_scr[index_close]))
print(rec_score)
```


...

[illegible]

