

```
In [1]: #Import Libraries
import os
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: #import data set, movies and credits
movies = pd.read_csv('movies.csv')
credits = pd.read_csv('credits.csv')
```

```
In [3]: movies.head(2)
```

```
Out[3]:
```

	budget	genres	homepage	id	keywords	original_lai
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1464, "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"}]	

```
In [4]: credits.head(2)
```

```
Out[4]:
```

	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 Sparrow", "credit_id": "52fe4232c3a36847f800b579", "de..."}]	

```
In [5]: print(movies.shape)
print()
print(credits.shape)
```

```
(4803, 20)
```

```
(4803, 4)
```

```
In [6]: print(movies.info())
print()
print(credits.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
None
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 4 columns):
#   Column    Non-Null Count  Dtype
---  -
0   movie_id  4803 non-null  int64
1   title     4803 non-null  object
2   cast      4803 non-null  object
3   crew      4803 non-null  object
dtypes: int64(1), object(3)
memory usage: 150.2+ KB
None
```

```
In [7]: print(movies.columns)
print()
print(credits.columns)
```

```
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')

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

```
In [8]: # Join both the table

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



In [12]:

#delete unwanted columns

movies =

movies[['movie\_id','title','overview','genres','keywords','cast', 'c

In [13]:

movies.head()

Out[13]:

	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", "...	[{"credi...
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...	[{"credi...
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...	[{"credi...
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...	[{"credi...
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...	[{"credi...

In [14]:

movies['genres'][0]

Out[14]:

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

In [15]:

movies['genres'][1]

Out[15]:

'[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 28, "name": "Action"}]'

```
In [16]: # check missing data
movies.isnull().sum()
```

```
Out[16]: movie_id    0
         title      0
         overview    3
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [17]: #drop null values
movies.dropna(inplace=True)
```

```
In [18]: movies.isnull().sum()
```

```
Out[18]: movie_id    0
         title      0
         overview    0
         genres      0
         keywords    0
         cast        0
         crew        0
         dtype: int64
```

```
In [19]: # To extract details from dictionaries - ast
import ast
```

```
In [20]:
def convert(text):
    L = []
    for i in ast.literal_eval(text):
        L.append(i['name'])
    return L
```

```
In [21]: movies.head(2)
```

```
Out[21]:
```

	movie_id	title	overview	genres	keywords	cast	c
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", "...	[{"credit...
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...	[{"credit...

```
In [22]: movies['genres'] = movies['genres'].apply(convert)
```

```
In [23]: movies.head()
```

```
Out[23]:
```

	movie_id	title	overview	genres	keywords	cast	cr
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", "...	{{"credit_...
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...	{{"credit_...
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...	{{"credit_...
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...	{{"credit_...
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...	{{"credit_...

```
In [24]: movies['keywords'][0]
```

```
Out[24]: '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "society"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"id": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "alien"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien planet"}, {"id": 10987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldier"}, {"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165431, "name": "anti war"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "name": "mind and soul"}, {"id": 209714, "name": "3d"}]'
```

```
In [25]: movies['keywords'] = movies['keywords'].apply(convert)
```

```
In [26]: movies.head()
```

Out[26]:

	movie_id	title	overview	genres	keywords	cast	cr
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", "...	[{"credit_": "52fe48009251416c750aca2"
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...	[{"credit_": "52fe4232c3a36847f800b57"
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...	[{"credit_": "54805967c3a36829b5002c4"
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...	[{"credit_": "52fe4781c3a36847f81398c"
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...	[{"credit_": "52fe479ac3a36847f813ea2"



```
In [27]: movies['cast'][0]
```

localhost:8888/notebooks/project/Project Session track- May 2023-20231106T043400Z-001/Raghav k Projects/Module8- 27 jan2024- recommenda... 10/24

```

st_id": 49, "character": "Link Room Tech", "credit_id": "52fe48009251416c750
aca9b", "gender": 0, "id": 1207247, "name": "Sonia Yee", "order": 28}, {"cas
t_id": 50, "character": "Basketball Avatar / Troupe", "credit_id": "52fe4800
9251416c750aca9f", "gender": 1, "id": 1207248, "name": "Jahnel Curfman", "or
der": 29}, {"cast_id": 51, "character": "Basketball Avatar", "credit_id": "5
2fe48009251416c750acaa3", "gender": 0, "id": 89714, "name": "Ilram Choi", "o
rder": 30}, {"cast_id": 52, "character": "Na\\'vi Child", "credit_id": "52fe4
8009251416c750acaa7", "gender": 0, "id": 1207249, "name": "Kyla Warren", "or
der": 31}, {"cast_id": 53, "character": "Troupe", "credit_id": "52fe48009251
416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Roumain", "order":
32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52fe48009251416c75
0acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"c
ast_id": 57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb",
"gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id":
55, "character": "Troupe", "credit_id": "52fe48009251416c750acab3", "gende
r": 0, "id": 1207251, "name": "Taylor Kibby", "order": 35}, {"cast_id": 56,
"character": "Troupe", "credit_id": "52fe48009251416c750acab7", "gender": 0,
"id": 1207252, "name": "Jodie Landau", "order": 36}, {"cast_id": 58, "charac
ter": "Troupe", "credit_id": "52fe48009251416c750acabf", "gender": 0, "id":
1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "character": "T
roupe", "credit_id": "52fe48009251416c750acac3", "gender": 0, "id": 1207257,
"name": "Cullen B. Madden", "order": 38}, {"cast_id": 60, "character": "Trou
pe", "credit_id": "52fe48009251416c750acac7", "gender": 0, "id": 1207259, "n
ame": "Joseph Brady Madden", "order": 39}, {"cast_id": 61, "character": "Tro
upe", "credit_id": "52fe48009251416c750acacb", "gender": 0, "id": 1207262,
"name": "Frankie Torres", "order": 40}, {"cast_id": 62, "character": "Troup
e", "credit_id": "52fe48009251416c750acacf", "gender": 1, "id": 1158600, "na
me": "Austin Wilson", "order": 41}, {"cast_id": 63, "character": "Troupe",
"credit_id": "52fe48019251416c750acad3", "gender": 1, "id": 983705, "name":
"Sara Wilson", "order": 42}, {"cast_id": 64, "character": "Troupe", "credit_
id": "52fe48019251416c750acad7", "gender": 0, "id": 1207263, "name": "Tamica
Washington-Miller", "order": 43}, {"cast_id": 65, "character": "Op Center St
aff", "credit_id": "52fe48019251416c750acadb", "gender": 1, "id": 1145098,
"name": "Lucy Briant", "order": 44}, {"cast_id": 66, "character": "Op Center
Staff", "credit_id": "52fe48019251416c750acadf", "gender": 2, "id": 33305,
"name": "Nathan Meister", "order": 45}, {"cast_id": 67, "character": "Op Cen
ter Staff", "credit_id": "52fe48019251416c750acae3", "gender": 0, "id": 1207
264, "name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op C
enter Staff", "credit_id": "52fe48019251416c750acae7", "gender": 2, "id": 33
311, "name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "characte
r": "Op Center Staff", "credit_id": "52fe48019251416c750acae8", "gender": 0,
"id": 1207265, "name": "Paul Yates", "order": 48}, {"cast_id": 70, "characte
r": "Op Center Duty Officer", "credit_id": "52fe48019251416c750acaef", "gend
er": 0, "id": 1207266, "name": "Wray Wilson", "order": 49}, {"cast_id": 71,
"character": "Op Center Staff", "credit_id": "52fe48019251416c750acaf3", "ge
nder": 2, "id": 54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72,
"character": "Dancer", "credit_id": "52fe48019251416c750acaf7", "gender": 0,
"id": 1207267, "name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 7
3, "character": "Dancer", "credit_id": "52fe48019251416c750acafb", "gender":
0, "id": 1207268, "name": "Carvon Futrell", "order": 52}, {"cast_id": 74, "c
haracter": "Dancer", "credit_id": "52fe48019251416c750acaff", "gender": 0,
"id": 1207269, "name": "Brandon Jelkes", "order": 53}, {"cast_id": 75, "char
acter": "Dancer", "credit_id": "52fe48019251416c750acb03", "gender": 0, "i
d": 1207270, "name": "Micah Moch", "order": 54}, {"cast_id": 76, "characte
r": "Dancer", "credit_id": "52fe48019251416c750acb07", "gender": 0, "id": 12
07271, "name": "Hanniyah Muhammad", "order": 55}, {"cast_id": 77, "characte
r": "Dancer", "credit_id": "52fe48019251416c750acb0b", "gender": 0, "id": 12

```

```

07272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "credit_id": "52fe48019251416c750acb0f", "gender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_id": 79, "character": "Dancer", "credit_id": "52fe48019251416c750acb13", "gender": 0, "id": 1207274, "name": "April Marie Thomas", "order": 58}, {"cast_id": 80, "character": "Dancer", "credit_id": "52fe48019251416c750acb17", "gender": 0, "id": 1207275, "name": "Bravita A. Threatt", "order": 59}, {"cast_id": 81, "character": "Mining Chief (uncredited)", "credit_id": "52fe48019251416c750acb1b", "gender": 0, "id": 1207276, "name": "Colin Bleasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (uncredited)", "credit_id": "52fe48019251416c750acb1f", "gender": 0, "id": 107969, "name": "Mike Bodnar", "order": 61}, {"cast_id": 83, "character": "Richard (uncredited)", "credit_id": "52fe48019251416c750acb23", "gender": 0, "id": 1207278, "name": "Matt Clayton", "order": 62}, {"cast_id": 84, "character": "Nav'i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gender": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "order": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2, "id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 89, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 90, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 91, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 92, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 93, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 94, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 95, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 96, "character": "Ops Centreworker (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 97, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 98, "character": "Col. Quaritch's Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "T. J. Storm", "order": 77}, {"cast_id": 99, "character": "Female Marine (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name": "Jodie Taylor", "order": 78}, {"cast_id": 100, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 101, "character": "Geologist (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207292, "name": "Richard Whiteside", "order": 80}, {"cast_id": 102, "character": "Na'vi (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Niki Zambo", "order": 81}, {"cast_id": 103, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}]

```

```
In [28]: 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 [29]: movies['cast'] = movies['cast'].apply(convert3)
```

```
In [30]: movies.head()
```

Out[30]:

	movie_id	title	overview	genres	keywords	cast	c
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 Weaver]	{{"credit
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]	{{"credit
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]	{{"credit
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]	{{"credit
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]	{{"credit

In [31]: `movies['crew'][0]`

Out[31]: `'[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": "Editor", "name": "Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gender": 0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "539c4a4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "5544ee3b925141499f0008fc", "department": "Sound", "gender": 2, "id": 1729, "job": "Original Music Composer", "name": "James Horner"}, {"credit_id": "52fe48009251416c750ac9c3", "department": "Directing", "gender": 2, "id": 2710, "job": "Director", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "gender": 2, "id": 2710, "job": "Writer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "department": "Editing", "gender": 2, "id": 2710, "job": "Editor", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca29", "department": "Production", "gender": 2, "id": 2710, "job": "Produce"}]`

In [32]: `def fetch_director(text):  
 L = []  
 for i in ast.literal_eval(text):  
 if i['job']=='Director':  
 L.append(i['name'])  
 return L`

In [33]: `movies['crew'] = movies['crew'].apply(fetch_director)`

In [34]: `movies.head()`

Out[34]:

	movie_id	title	overview	genres	keywords	cast	crew
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 Weaver]	[James Cameron]
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]	[Gore Verbinski]
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]	[Sam Mendes]
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]	[Christopher Nolan]
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]	[Andrew Stanton]

```
In [35]: def collapse(L):
          L1 = []
          for i in L:
              L1.append(i.replace(" ", ""))
          return L1
```

In [36]: `movies.columns`

Out[36]: Index(['movie\_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew'], dtype='object')

```
In [37]: movies['genres'] = movies['genres'].apply(collapse)
          movies['keywords'] = movies['keywords'].apply(collapse)
          movies['cast'] = movies['cast'].apply(collapse)
          movies['crew'] = movies['crew'].apply(collapse)
```

```
In [38]: movies.head()
```

Out[38]:

	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 [39]: print(movies['overview'][0])
print("*****" * 10)
print(movies['genres'][0])
print("*****" * 10)
print(movies['keywords'][0])
print("*****" * 10)
print(movies['cast'][0])
print("*****" * 10)
print(movies['crew'][0])
print("*****" * 10)
```

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 an alien civilization.

```
*****
****
['Action', 'Adventure', 'Fantasy', 'ScienceFiction']
*****
****
['cultureclash', 'future', 'spacewar', 'spacecolony', 'society', 'spacetravel', 'futuristic', 'romance', 'space', 'alien', 'tribe', 'alienplanet', 'cgi', 'marine', 'soldier', 'battle', 'loveaffair', 'antiwar', 'powerrelations', 'mindandsoul', '3d']
*****
****
['SamWorthington', 'ZoeSaldana', 'SigourneyWeaver']
*****
****
['JamesCameron']
*****
****
```

```
In [40]: # split sentence into words
movies['overview'] = movies['overview'].apply(lambda x: x.split())
```

```
In [41]: print(movies['overview'][0])
```

```
['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', 'an', 'alien', 'civilization.']
```

```
In [42]: print(movies['overview'][0])
print("*****" * 10)
print(movies['genres'][0])
print("*****" * 10)
print(movies['keywords'][0])
print("*****" * 10)
print(movies['cast'][0])
print("*****" * 10)
print(movies['crew'][0])
print("*****" * 10)
```

```
['In', 'the', '22nd', 'century,', 'a', 'paraplegic', 'Marine', 'is', 'dispat
ched', 'to', 'the', 'moon', 'Pandora', 'on', 'a', 'unique', 'mission,', 'bu
t', 'becomes', 'torn', 'between', 'following', 'orders', 'and', 'protectin
g', 'an', 'alien', 'civilization.']
```

```
*****
****
```

```
['Action', 'Adventure', 'Fantasy', 'ScienceFiction']
```

```
*****
****
```

```
['cultureclash', 'future', 'spacewar', 'spacecolony', 'society', 'spacetrave
l', 'futuristic', 'romance', 'space', 'alien', 'tribe', 'alienplanet', 'cg
i', 'marine', 'soldier', 'battle', 'loveaffair', 'antiwar', 'powerrelation
s', 'mindandsoul', '3d']
```

```
*****
****
```

```
['SamWorthington', 'ZoeSaldana', 'SigourneyWeaver']
```

```
*****
****
```

```
['JamesCameron']
```

```
*****
****
```

```
In [43]: # make one columns basis overview, genres, keywords, cast and crew

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

In [44]: `movies.head()`

Out[44]:

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

In [45]: `new = movies.drop(columns=['overview', 'genres', 'keywords', 'cast', 'crew'])`

In [46]: `new.head()`

Out[46]:

	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...

In [47]: `new['tags'] = new['tags'].apply(lambda x: " ".join(x))`

In [48]: `new.head()`

Out[48]:

	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 [49]: `new['tags'][0]`

Out[49]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protectin  
g an alien civilization. Action Adventure Fantasy ScienceFiction cultureclas  
h future spacewar spacecolony society spacetravel futuristic romance space a  
lien tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrel  
ations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCamero  
n'

## Bag of words (BOW) - Feature extraction/engineering method

In [50]: `from sklearn.feature_extraction.text import CountVectorizer  
# apply stopwords and also give hard coded feature value - 5000  
cv = CountVectorizer(stop_words='english', max_features=5000)  
vector = cv.fit_transform(new['tags']).toarray()  
vector.shape`

Out[50]: (4806, 5000)

## To use Similarity matrix for building recommendation system

In [51]: `from sklearn.metrics.pairwise import cosine_similarity`

In [52]: `similarity = cosine_similarity(vector)`

```
In [53]: similarity
```

```
Out[53]: 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 [54]: #near by 1 is similar movie
```

```
In [55]: new.tail(10)
```

```
Out[55]:
```

	movie_id	title	tags
<b>4799</b>	182291	On The Downlow	Isaac and Angel are two young Latinos involved...
<b>4800</b>	286939	Sanctuary: Quite a Conundrum	It should have been just a normal day of sex, ...
<b>4801</b>	124606	Bang	A young woman in L.A. is having a bad day: she...
<b>4802</b>	14337	Primer	Friends/fledgling entrepreneurs invent a devic...
<b>4803</b>	67238	Cavite	Adam, a security guard, travels from Californi...
<b>4804</b>	9367	El Mariachi	El Mariachi just wants to play his guitar and ...
<b>4805</b>	72766	Newlyweds	A newlywed couple's honeymoon is upended by th...
<b>4806</b>	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...
<b>4807</b>	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...
<b>4808</b>	25975	My Date with Drew	Ever since the second grade when he first saw ...

```
In [56]: new[new['title']=='Bang'].index[0]
```

```
Out[56]: 4801
```

## Recommendation System

```
In [57]: def recommend(movie):  
        index = new[new['title']==movie].index[0]  
        distances = sorted(list(enumerate(similarity[index])), reverse=True, key=  
  
        for i in distances[1:6]:  
            print(new.iloc[i[0]].title)
```

```
In [58]: recommend('Spider-Man 3')
```

Spider-Man 2  
Spider-Man  
The Amazing Spider-Man 2  
The Amazing Spider-Man  
Arachnophobia

```
In [60]: recommend('Harry Potter and the Chamber of Secrets')
```

Harry Potter and the Prisoner of Azkaban  
Harry Potter and the Order of the Phoenix  
Harry Potter and the Goblet of Fire  
Harry Potter and the Philosopher's Stone  
Harry Potter and the Half-Blood Prince

```
In [61]: recommend('The Dark Knight Rises')
```

The Dark Knight  
Batman Begins  
Batman  
Batman Returns  
Batman

```
In [62]: recommend('Avengers: Age of Ultron')
```

Iron Man  
Iron Man 3  
The Avengers  
Captain America: Civil War  
Iron Man 2

```
In [64]: recommend('Primer')
```

Purple Violets  
She's the One  
My Big Fat Greek Wedding 2  
Lovely & Amazing  
I Don't Know How She Does It

```
In [ ]:
```

## Deployment

```
In [65]: import pickle
```

```
In [68]: !pip install streamlit
```

Collecting streamlit

WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)  
WARNING: Ignoring invalid distribution -ordcloud (c:\users\abc\anaconda3\lib\site-packages)

```
In [66]: pickle.dump(new.to_dict(), open('movie_dict.pkl', 'wb'))  
pickle.dump(similarity, open('similarity.pkl', 'wb'))
```

```
In [ ]: import pickle
import streamlit as st
import pandas as pd

def recommend(movie):
    movie_index = movies[movies['title']==movie].index[0]
    distances = similarity[movie_index]
    movies_list = sorted(list(enumerate(distances)), reverse=True, key= lambda
    recommended_movies = []
    for i in movies_list:
        recommended_movies.append(movies.iloc[i[0]].title)
    return recommended_movies

movies_dict = pickle.load(open('movie_dict.pkl', 'rb'))
movies = pd.DataFrame(movies_dict)
similarity = pickle.load(open('similarity.pkl', 'rb'))

st.title("Movie Recommender System-Raghav")

selected_movie_name = st.selectbox(
    'How would you like to be contacted ?',
    movies['title'].values)

if st.button('Recommend'):
    recommendations = recommend(selected_movie_name)

for i in recommendations:
    st.write(i)
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: