

```
In [1]: import numpy as np  
import pandas as pd
```

```
In [8]: movies=pd.read_csv("tmdb_5000_movies.csv")
```

```
In [3]: credits=pd.read_csv("tmdb_5000_credits.csv")
```

```
In [9]: movies.head(1)
```

```
Out[9]:
```

	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	



```
In [10]: credits.head(1)
```

```
Out[10]:
```

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...]	[{"credit_id": "52fe48009251416c750aca23", "de...]

```
In [11]: movies = movie.merge(credits, on='title')  
movies.shape
```

```
Out[11]: (4809, 23)
```

```
In [12]: movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4809 entries, 0 to 4808
Data columns (total 23 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   budget            4809 non-null   int64  
 1   genres             4809 non-null   object  
 2   homepage          1713 non-null   object  
 3   id                4809 non-null   int64  
 4   keywords          4809 non-null   object  
 5   original_language 4809 non-null   object  
 6   original_title    4809 non-null   object  
 7   overview          4806 non-null   object  
 8   popularity         4809 non-null   float64 
 9   production_companies 4809 non-null   object  
 10  production_countries 4809 non-null   object  
 11  release_date      4808 non-null   object  
 12  revenue            4809 non-null   int64  
 13  runtime            4807 non-null   float64 
 14  spoken_languages   4809 non-null   object  
 15  status              4809 non-null   object  
 16  tagline            3965 non-null   object  
 17  title               4809 non-null   object  
 18  vote_average       4809 non-null   float64 
 19  vote_count          4809 non-null   int64  
 20  movie_id           4809 non-null   int64  
 21  cast                4809 non-null   object  
 22  crew                4809 non-null   object  
dtypes: float64(3), int64(5), object(15)
memory usage: 864.2+ KB
```

```
In [13]: # which Columns we need?
```

```
# genres
# id
# keywords
# overview
# title
# cast
# crew
```

```
In [14]: movies=movies[['movie_id','title','overview','keywords','genres','cast','crew']]
```

```
In [15]: movies.head()
```

	movie_id	title	overview	keywords	genres	cast	
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 1463, "name": "culture clash"}, {"id": ...]	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	{"character": "Jake Sully", "name": "Sully", "id": 242, "cast_id": 28, "order": 1}, {"character": "Neelix", "name": "Neelix", "id": 14, "cast_id": 12, "order": 2}, {"character": "Miles", "name": "Miles", "id": 12, "cast_id": 28, "order": 3}, {"character": "Tenzin", "name": "Tenzin", "id": 11, "cast_id": 12, "order": 4}, {"character": "Moat", "name": "Moat", "id": 10, "cast_id": 12, "order": 5}, {"character": "Oli", "name": "Oli", "id": 9, "cast_id": 12, "order": 6}, {"character": "Pakku", "name": "Pakku", "id": 8, "cast_id": 12, "order": 7}, {"character": "Raya", "name": "Raya", "id": 7, "cast_id": 12, "order": 8}, {"character": "Tuktuk", "name": "Tuktuk", "id": 6, "cast_id": 12, "order": 9}, {"character": "Viper", "name": "Viper", "id": 5, "cast_id": 12, "order": 10}, {"character": "Wol", "name": "Wol", "id": 4, "cast_id": 12, "order": 11}, {"character": "Zuko", "name": "Zuko", "id": 3, "cast_id": 12, "order": 12}, {"character": "Zut", "name": "Zut", "id": 2, "cast_id": 12, "order": 13}, {"character": "Zut", "name": "Zut", "id": 1, "cast_id": 12, "order": 14}	{"crec": "52fe48009251416c750a"}
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "ha..."]]	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}]	{"character": "Jack Sparrow", "name": "Jack Sparrow", "id": 4, "cast_id": 12, "order": 1}, {"character": "Will Turner", "name": "Will Turner", "id": 11, "cast_id": 12, "order": 2}, {"character": "Cotton", "name": "Cotton", "id": 10, "cast_id": 12, "order": 3}, {"character": "Bootstrap Bill Turner", "name": "Bootstrap Bill Turner", "id": 9, "cast_id": 12, "order": 4}, {"character": "Marty", "name": "Marty", "id": 8, "cast_id": 12, "order": 5}, {"character": "Bootstrap", "name": "Bootstrap", "id": 7, "cast_id": 12, "order": 6}, {"character": "Tia Dalma", "name": "Tia Dalma", "id": 6, "cast_id": 12, "order": 7}, {"character": "Marty", "name": "Marty", "id": 5, "cast_id": 12, "order": 8}, {"character": "Bootstrap", "name": "Bootstrap", "id": 4, "cast_id": 12, "order": 9}, {"character": "Marty", "name": "Marty", "id": 3, "cast_id": 12, "order": 10}, {"character": "Bootstrap", "name": "Bootstrap", "id": 2, "cast_id": 12, "order": 11}, {"character": "Marty", "name": "Marty", "id": 1, "cast_id": 12, "order": 12}	{"crec": "52fe4232c3a36847f8001"}
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 470, "name": "spy"}, {"id": 818, "name": "name..."]]	[{"id": 28, "name": "Action"}, {"id": 12, "name": "James Bond"}]	{"character": "James Bond", "name": "James Bond", "id": 1, "cast_id": 28, "order": 1}, {"character": "M", "name": "M", "id": 10, "cast_id": 28, "order": 2}, {"character": "Q", "name": "Q", "id": 9, "cast_id": 28, "order": 3}, {"character": "Rami", "name": "Rami", "id": 8, "cast_id": 28, "order": 4}, {"character": "Monsieur", "name": "Monsieur", "id": 7, "cast_id": 28, "order": 5}, {"character": "Kingsley", "name": "Kingsley", "id": 6, "cast_id": 28, "order": 6}, {"character": "M", "name": "M", "id": 5, "cast_id": 28, "order": 7}, {"character": "Q", "name": "Q", "id": 4, "cast_id": 28, "order": 8}, {"character": "Rami", "name": "Rami", "id": 3, "cast_id": 28, "order": 9}, {"character": "Monsieur", "name": "Monsieur", "id": 2, "cast_id": 28, "order": 10}, {"character": "Kingsley", "name": "Kingsley", "id": 1, "cast_id": 28, "order": 11}	{"crec": "54805967c3a36829b500"}
3	49026	The Dark Knight Rises	Following the death of District Attorney Harvey...	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "..."]]	[{"id": 28, "name": "Action"}, {"id": 80, "name": "name..."]]	{"character": "Bruce Wayne / Batman", "name": "Bruce Wayne / Batman", "id": 2, "cast_id": 28, "order": 1}, {"character": "Selina Kyle", "name": "Selina Kyle", "id": 11, "cast_id": 28, "order": 2}, {"character": "Alfred Pennyworth", "name": "Alfred Pennyworth", "id": 10, "cast_id": 28, "order": 3}, {"character": "Commissioner Gordon", "name": "Commissioner Gordon", "id": 9, "cast_id": 28, "order": 4}, {"character": "Harvey Dent", "name": "Harvey Dent", "id": 8, "cast_id": 28, "order": 5}, {"character": "Lucius Fox", "name": "Lucius Fox", "id": 7, "cast_id": 28, "order": 6}, {"character": "Dr. Jonathan Crane", "name": "Dr. Jonathan Crane", "id": 6, "cast_id": 28, "order": 7}, {"character": "Lucius Fox", "name": "Lucius Fox", "id": 5, "cast_id": 28, "order": 8}, {"character": "Dr. Jonathan Crane", "name": "Dr. Jonathan Crane", "id": 4, "cast_id": 28, "order": 9}, {"character": "Lucius Fox", "name": "Lucius Fox", "id": 3, "cast_id": 28, "order": 10}, {"character": "Dr. Jonathan Crane", "name": "Dr. Jonathan Crane", "id": 2, "cast_id": 28, "order": 11}, {"character": "Lucius Fox", "name": "Lucius Fox", "id": 1, "cast_id": 28, "order": 12}	{"crec": "52fe4781c3a36847f813"}
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 818, "name": "based on novel"}, {"id": ...}]	[{"id": 28, "name": "Action"}, {"id": 12, "name": "name..."]]	{"character": "John Carter", "name": "John Carter", "id": 5, "cast_id": 28, "order": 1}, {"character": "Dejah Thoris", "name": "Dejah Thoris", "id": 11, "cast_id": 28, "order": 2}, {"character": "Tars Tarkas", "name": "Tars Tarkas", "id": 10, "cast_id": 28, "order": 3}, {"character": "Aeolus", "name": "Aeolus", "id": 9, "cast_id": 28, "order": 4}, {"character": "Thark", "name": "Thark", "id": 8, "cast_id": 28, "order": 5}, {"character": "Aeolus", "name": "Aeolus", "id": 7, "cast_id": 28, "order": 6}, {"character": "Thark", "name": "Thark", "id": 6, "cast_id": 28, "order": 7}, {"character": "Aeolus", "name": "Aeolus", "id": 5, "cast_id": 28, "order": 8}, {"character": "Thark", "name": "Thark", "id": 4, "cast_id": 28, "order": 9}, {"character": "Aeolus", "name": "Aeolus", "id": 3, "cast_id": 28, "order": 10}, {"character": "Thark", "name": "Thark", "id": 2, "cast_id": 28, "order": 11}, {"character": "Aeolus", "name": "Aeolus", "id": 1, "cast_id": 28, "order": 12}	{"crec": "52fe479ac3a36847f813"}

In [16]: `movies.isnull().sum()`

Out[16]:

movie_id	0
title	0
overview	3
keywords	0
genres	0
cast	0
crew	0
dtype: int64	

In [17]: `movies.dropna(inplace=True)`

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

Out[18]: 0

```
In [19]: movies.iloc[0].genres
```

```
Out[19]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [20]: import ast  
def converter(obj):  
    l = []  
    for i in ast.literal_eval(obj):  
        l.append(i['name'])  
    return l
```

```
In [26]: movies['genres']=movies['genres'].apply(converter)
```

```
In [29]: movies.head(2)
```

```
Out[29]:
```

	movie_id	title	overview	keywords	genres	cast	credit_id
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 1463, "name": "culture clash"}, {"id": ...]	[Action, Adventure, Fantasy, Science Fiction]	[{"cast_id": 242, "character": "Jake Sully", "..."]	[{"credit_id": "52fe48009251416c750aca...", "order": 1}
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "..."]	[Adventure, Fantasy, Action]	[{"cast_id": 4, "character": "Captain Jack Spa..."]	[{"credit_id": "52fe4232c3a36847f800b5...", "order": 2}]

```
In [30]: movies['keywords']=movies['keywords'].apply(converter)
```

```
In [31]: movies.head(2)
```

```
Out[31]:
```

	movie_id	title	overview	keywords	genres	cast	credit_id
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[culture clash, future, space war, space colon...]	[Action, Adventure, Fantasy, Science Fiction]	[{"cast_id": 242, "character": "Jake Sully", "..."]	[{"credit_id": "52fe48009251416c750aca...", "order": 1}
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[ocean, drug abuse, exotic island, east india ...]	[Adventure, Fantasy, Action]	[{"cast_id": 4, "character": "Captain Jack Spa..."]	[{"credit_id": "52fe4232c3a36847f800b5...", "order": 2}]

```
In [32]: def converter2(obj):
    l = []
    counter = 0
    for i in ast.literal_eval(obj):
        if counter != 3:
            l.append(i['name'])
            counter+=1
        else:
            break
    return l
```

```
In [34]: movies['cast']=movies['cast'].apply(converter2)
```

```
In [35]: movies.head(2)
```

Out[35]:

	movie_id	title	overview	keywords	genres	cast
0	1995	Avatar	In the 22nd century, a paraplegic Marine is di...	[culture clash, future, space war, space colon...]	[Action, Adventure, Fantasy, Science Fiction]	[Sam Worthington, Zoe Saldana, Sigourney Weaver] [{"credi
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[ocean, drug abuse, exotic island, east india ...]	[Adventure, Fantasy, Action]	[Johnny Depp, Orlando Bloom, Keira Knightley] [{"credi

```
In [37]: movies['crew'][0]
```

```
Out[37]: '[{"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": "Producer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca3f", "department": "Writing", "gender": 2, "id": 2710, "job": "Screenplay", "name": "James Cameron"}, {"credit_id": "539c4987c3a36810ba0021a4", "department": "Art", "gender": 2, "id": 7236, "job": "Art Direction", "name": "Andrew Menzies"}, {"credit_id": "549598c3c3a3686ae9004383", "department": "Visual Effects", "gender": 0, "id": 6690, "job": "Visual Effects Producer", "name": "Jill Brooks"}, {"credit_id": "52fe48009251416c750aca4b", "department": "Production", "gender": 1, "id": 6347, "job": "Casting", "name": "Margery Simkin"}, {"credit_id": "570b6f419251417da70032fe", "department": "Art", "gender": 2, "id": 6878, "job": "Supervising Art Director", "name": "Kevin Ishioka"}, {"credit_id": "5495a0fac3a3686ae9004468", "department": "Sound", "gender": 0, "id": 6883, "job": "Music Editor", "name": "Dick Bernstein"}, {"credit_id": "54959706c3a3686af3003e81", "department": "Sound", "gender": 0, "id": 8159, "job": "Sound Effects Editor", "name": "Shannon Mills"}, {"credit_id": "54491d58c3a3680fb1001ccb", "department": "Sound", "gender": 0, "id": 8160, "job": "Foley", "name": "Dennie Thorpe"}, {"credit_id": "54491d6cc3a3680fa5001b2c", "department": "Sound", "gender": 0, "id": 8163, "job": "Foley", "name": "Jana Vance"}, {"credit_id": "52fe48009251416c750aca57", "department": "Costume & Make-Up", "gender": 1, "id": 8527, "job": "Costume Design", "name": "Deborah Lynn Scott"}, {"credit_id": "52fe48009251416c750aca2f", "department": "Production", "gender": 2, "id": 8529, "job": "Producer", "name": "Jon Landau"}, {"credit_id": "539c4937c3a36810ba002194", "department": "Art", "gender": 0, "id": 9618, "job": "Art Direction", "name": "Sean Haworth"}, {"credit_id": "539c49b6c3a36810c10020e6", "department": "Art", "gender": 1, "id": 12653, "job": "Set Decoration", "name": "Kim Sinclair"}, {"credit_id": "570b6f2f9251413a0e00020d", "department": "Art", "gender": 1, "id": 12653, "job": "Supervising Art Director", "name": "Kim Sinclair"}, {"credit_id": "54491a6c0e0a26748c001b19", "department": "Art", "gender": 2, "id": 14350, "job": "Set Designer", "name": "Richard F. Mays"}, {"credit_id": "56928cf4c3a3684cff0025c4", "department": "Production", "gender": 1, "id": 20294, "job": "Executive Producer", "name": "Laeta Kalogridis"}, {"credit_id": "52fe48009251416c750aca51", "department": "Costume & Make-Up", "gender": 0, "id": 17675, "job": "Costume Design", "name": "Mayes C. Rubeo"}, {"credit_id": "52fe48009251416c750aca11", "department": "Camera", "gender": 2, "id": 18265, "job": "Director of Photography", "name": "Mauro Fiore"}, {"credit_id": "5449194d0e0a26748f001b39", "department": "Art", "gender": 0, "id": 42281, "job": "Set Designer", "name": "Scott Herbertson"}, {"credit_id": "52fe48009251416c750aca05", "department": "Crew", "gender": 0, "id": 42288, "job": "Stunts", "name": "Woody Schultz"}, {"credit_id": "5592aefb92514152de0010f5", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Makeup Artist", "name": "Linda DeVetta"}, {"credit_id": "5592afa492514152de00112c", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Hairstylist", "name": "Linda DeVett"}, {"credit_id": "54959ed592514130fc002e5d", "department": "Camera", "gender": 2, "id": 33302, "job": "Camera Operator", "name": "Richard Bluck"}, {"credit_id": "539c4891c3a36810ba002147", "department": "Art", "gender": 2, "id": 33303, "job": "Art Direction", "name": "Simon Bright"}, {"credit_id": "54959c069251417a81001f3a", "department": "Visual Effects", "gender": 0, "id": 113145, "job": "Visual Effects Supervisor", "name": "Richard Martin"}, {"credit_id": "54959a0dc3a3680ff5002c8d", "department": "C
```

rew", "gender": 2, "id": 58188, "job": "Visual Effects Editor", "name": "Steve R. Moore"}, {"credit\_id": "52fe48009251416c750aca1d", "department": "Editing", "gender": 2, "id": 58871, "job": "Editor", "name": "John Refoua"}, {"credit\_id": "54491a4dc3a3680fc30018ca", "department": "Art", "gender": 0, "id": 92359, "job": "Set Designer", "name": "Karl J. Martin"}, {"credit\_id": "52fe48009251416c750aca35", "department": "Catering", "gender": 1, "id": 72201, "job": "Director of Photography", "name": "Chiling Lin"}, {"credit\_id": "52fe48009251416c750ac9ff", "department": "Crew", "gender": 0, "id": 89714, "job": "Stunts", "name": "Ilram Choi"}, {"credit\_id": "54959c529251416e2b004394", "department": "Visual Effects", "gender": 2, "id": 93214, "job": "Visual Effects Supervisor", "name": "Steven Quale"}, {"credit\_id": "54491edf0e0a267489001c37", "department": "Crew", "gender": 1, "id": 122607, "job": "Dialect Coach", "name": "Carla Meyer"}, {"credit\_id": "539c485bc3a368653d001a3a", "department": "Art", "gender": 2, "id": 132585, "job": "Art Direction", "name": "Nick Bassett"}, {"credit\_id": "539c4903c3a368653d001a74", "department": "Art", "gender": 0, "id": 132596, "job": "Art Direction", "name": "Jill Cormack"}, {"credit\_id": "539c4967c3a368653d001a94", "department": "Art", "gender": 0, "id": 132604, "job": "Art Direction", "name": "Andy McLaren"}, {"credit\_id": "52fe48009251416c750aca45", "department": "Crew", "gender": 0, "id": 236696, "job": "Motion Capture Artist", "name": "Terry Notary"}, {"credit\_id": "54959e02c3a3680fc60027d2", "department": "Crew", "gender": 2, "id": 956198, "job": "Stunt Coordinator", "name": "Garrett Warren"}, {"credit\_id": "54959ca3c3a3686ae300438c", "department": "Visual Effects", "gender": 2, "id": 957874, "job": "Visual Effects Supervisor", "name": "Jonathan Rothbart"}, {"credit\_id": "570b6f519251412c74001b2f", "department": "Art", "gender": 0, "id": 957889, "job": "Supervising Art Director", "name": "Stefan Dechant"}, {"credit\_id": "570b6f62c3a3680b77007460", "department": "Art", "gender": 2, "id": 959555, "job": "Supervising Art Director", "name": "Todd Cherniawsky"}, {"credit\_id": "539c4a3ac3a36810da0021cc", "department": "Production", "gender": 0, "id": 1016177, "job": "Casting", "name": "Miranda Rivers"}, {"credit\_id": "539c482cc3a36810c1002062", "department": "Art", "gender": 0, "id": 1032536, "job": "Production Design", "name": "Robert Stromberg"}, {"credit\_id": "539c4b65c3a36810c9002125", "department": "Costume & Make-Up", "gender": 2, "id": 1071680, "job": "Costume Design", "name": "John Harding"}, {"credit\_id": "54959e6692514130fc002e4e", "department": "Camera", "gender": 0, "id": 1177364, "job": "Steadicam Operator", "name": "Roberto De Angelis"}, {"credit\_id": "539c49f1c3a368653d001aac", "department": "Costume & Make-Up", "gender": 2, "id": 1202850, "job": "Makeup Department Head", "name": "Mike Smithson"}, {"credit\_id": "5495999ec3a3686ae100460c", "department": "Visual Effects", "gender": 0, "id": 1204668, "job": "Visual Effects Producer", "name": "Alain Lalanne"}, {"credit\_id": "54959cd9c3a3681153002729", "department": "Visual Effects", "gender": 0, "id": 1206410, "job": "Visual Effects Supervisor", "name": "Lucas Salton"}, {"credit\_id": "549596239251417a81001eae", "department": "Crew", "gender": 0, "id": 1234266, "job": "Post Production Supervisor", "name": "Janace Tashjian"}, {"credit\_id": "54959c859251416e1e003efe", "department": "Visual Effects", "gender": 0, "id": 1271932, "job": "Visual Effects Supervisor", "name": "Stephen Rosenbaum"}, {"credit\_id": "5592af28c3a368775a00105f", "department": "Costume & Make-Up", "gender": 0, "id": 1310064, "job": "Makeup Artist", "name": "Frankie Karena"}, {"credit\_id": "539c4adfc3a36810e300203b", "department": "Costume & Make-Up", "gender": 1, "id": 1319844, "job": "Costume Supervisor", "name": "Lisa Lovaas"}, {"credit\_id": "54959b579251416e2b004371", "department": "Visual Effects", "gender": 0, "id": 1327028, "job": "Visual Effects Supervisor", "name": "Jonathan Fawkner"}, {"credit\_id": "539c48a7c3a36810b5001fa7", "department": "Art", "gender": 0, "id": 1330561, "job": "Art Direction", "name": "Robert Bavin"}, {"credit\_id": "539c4a71c3a36810da0021e0", "department": "Costume & Make-Up", "gender": 0, "id": 1330567, "job": "Costume Supervisor", "name": "Anthony Almaraz"}, {"credit\_id": "539c4a8ac3a36810ba0021e4", "department": "Costume & Make-Up", "gender": 0, "id": 1330570, "job": "Costume Supervisor", "name": "Carolyn M. Fenton"}, {"credit\_id": "539c4ab6c3a36810da0021f0", "department": "Costume & Make-Up", "gender": 0, "id": 1330574, "job": "Costume Supervisor", "name": "Beth Koenigsberg"}, {"credit\_id": "54491ab70e0a267480001ba2", "department": "Art", "gender": 0, "id": 1336191, "job": "Set Designer", "name": "Sam Page"}, {"credit\_id": "544919d9c3a3680fc30018bd", "department": "Art", "gender": 0, "id": 1339441, "job": "Set Designer", "name": "Tex Kadonaga"}, {"credit\_id": "54491cf50e0a267483001b0c", "department": "Editing", "gender": 0, "id": 1352422, "job": "Dialogue Editor", "name": "Kim Foscato"}, {"credit\_id": "

"544919f40e0a26748c001b09", "department": "Art", "gender": 0, "id": 1352962, "job": "Set Designer", "name": "Tammy S. Lee"}, {"credit\_id": "5495a115c3a3680ff5002d71", "department": "Crew", "gender": 0, "id": 1357070, "job": "Transportation Coordinator", "name": "Denny Caira"}, {"credit\_id": "5495a12f92514130fc002e94", "department": "Crew", "gender": 0, "id": 1357071, "job": "Transportation Coordinator", "name": "James Waitkus"}, {"credit\_id": "5495976fc3a36811530026b0", "department": "Sound", "gender": 0, "id": 1360103, "job": "Supervising Sound Editor", "name": "Addison Teague"}, {"credit\_id": "54491837c3a3680fb1001c5a", "department": "Art", "gender": 2, "id": 1376887, "job": "Set Designer", "name": "C. Scott Baker"}, {"credit\_id": "54491878c3a3680fb4001c9d", "department": "Art", "gender": 0, "id": 1376888, "job": "Set Designer", "name": "Luke Caska"}, {"credit\_id": "544918dac3a3680fa5001ae0", "department": "Art", "gender": 0, "id": 1376889, "job": "Set Designer", "name": "David Chow"}, {"credit\_id": "544919110e0a267486001b68", "department": "Art", "gender": 0, "id": 1376890, "job": "Set Designer", "name": "Jonathan Dyer"}, {"credit\_id": "54491967c3a3680faa001b5e", "department": "Art", "gender": 0, "id": 1376891, "job": "Set Designer", "name": "Joseph Hiura"}, {"credit\_id": "54491997c3a3680fb1001c8a", "department": "Art", "gender": 0, "id": 1376892, "job": "Art Department Coordinator", "name": "Rebecca Jellie"}, {"credit\_id": "544919ba0e0a26748f001b42", "department": "Art", "gender": 0, "id": 1376893, "job": "Set Designer", "name": "Robert Andrew Johnson"}, {"credit\_id": "54491b1dc3a3680faa001b8c", "department": "Art", "gender": 0, "id": 1376895, "job": "Assistant Art Director", "name": "Mike Stassi"}, {"credit\_id": "54491b79c3a3680fbb001826", "department": "Art", "gender": 0, "id": 1376897, "job": "Construction Coordinator", "name": "John Villarino"}, {"credit\_id": "54491baec3a3680fb4001ce6", "department": "Art", "gender": 2, "id": 1376898, "job": "Assistant Art Director", "name": "Jeffrey Wisniewski"}, {"credit\_id": "54491d2fc3a3680fb4001d07", "department": "Editing", "gender": 0, "id": 1376899, "job": "Dialogue Editor", "name": "Cheryl Nardi"}, {"credit\_id": "54491d86c3a3680fa5001b2f", "department": "Editing", "gender": 0, "id": 1376901, "job": "Dialogue Editor", "name": "Marshall Winn"}, {"credit\_id": "54491d9dc3a3680faa001bb0", "department": "Sound", "gender": 0, "id": 1376902, "job": "Supervising Sound Editor", "name": "Gwendolyn Yates Whittle"}, {"credit\_id": "54491dc10e0a267486001bce", "department": "Sound", "gender": 0, "id": 1376903, "job": "Sound Re-Recording Mixer", "name": "William Stein"}, {"credit\_id": "54491f500e0a26747c001c07", "department": "Crew", "gender": 0, "id": 1376909, "job": "Choreographer", "name": "Lula Washington"}, {"credit\_id": "549599239251412c4e002a2e", "department": "Visual Effects", "gender": 0, "id": 1391692, "job": "Visual Effects Producer", "name": "Chris Del Conte"}, {"credit\_id": "54959d54c3a36831b8001d9a", "department": "Visual Effects", "gender": 2, "id": 1391695, "job": "Visual Effects Supervisor", "name": "R. Christopher White"}, {"credit\_id": "54959bdf9251412c4e002a66", "department": "Visual Effects", "gender": 0, "id": 1394070, "job": "Visual Effects Supervisor", "name": "Dan Lemmon"}, {"credit\_id": "5495971d92514132ed002922", "department": "Sound", "gender": 0, "id": 1394129, "job": "Sound Effects Editor", "name": "Tim Nielsen"}, {"credit\_id": "5592b25792514152cc001aa", "department": "Crew", "gender": 0, "id": 1394286, "job": "CG Supervisor", "name": "Michael Mulholland"}, {"credit\_id": "54959a329251416e2b004355", "department": "Crew", "gender": 0, "id": 1394750, "job": "Visual Effects Editor", "name": "Thomas Nittmann"}, {"credit\_id": "54959d6dc3a3686ae9004401", "department": "Visual Effects", "gender": 0, "id": 1394755, "job": "Visual Effects Supervisor", "name": "Edson Williams"}, {"credit\_id": "5495a08fc3a3686ae300441c", "department": "Editing", "gender": 0, "id": 1394953, "job": "Digital Intermediate", "name": "Christine Carr"}, {"credit\_id": "55402d659251413d6d000249", "department": "Visual Effects", "gender": 0, "id": 1395269, "job": "Visual Effects Supervisor", "name": "John Bruno"}, {"credit\_id": "54959e7b9251416e1e003f3e", "department": "Camera", "gender": 0, "id": 1398970, "job": "Steadicam Operator", "name": "David Emmerichs"}, {"credit\_id": "54959734c3a3686ae10045e0", "department": "Sound", "gender": 0, "id": 1400906, "job": "Sound Effects Editor", "name": "Christopher Scarabosio"}, {"credit\_id": "549595dd92514130fc002d79", "department": "Production", "gender": 0, "id": 1401784, "job": "Production Supervisor", "name": "Jennifer Teves"}, {"credit\_id": "549596009251413af70028cc", "department": "Production", "gender": 0, "id": 1401785, "job": "Production Manager", "name": "Bridgette Yorke"}, {"credit\_id": "549596e892514130fc002d99", "department": "Sound", "gender": 0, "id": 1401786, "job": "Sound Effects Editor", "name": "Ken Fischer"}, {"credit\_id": "549598229251412c4e002a1c", "department": "Crew", "gender": 0, "id": 1401787}

7, "job": "Special Effects Coordinator", "name": "Iain Hutton"}, {"credit\_id": "549598349251416e2b00432b", "department": "Crew", "gender": 0, "id": 1401788, "job": "Special Effects Coordinator", "name": "Steve Ingram"}, {"credit\_id": "54959905c3a3686ae3004324", "department": "Visual Effects", "gender": 0, "id": 1401789, "job": "Visual Effects Producer", "name": "Joyce Cox"}, {"credit\_id": "5495994b92514132ed002951", "department": "Visual Effects", "gender": 0, "id": 1401790, "job": "Visual Effects Producer", "name": "Jenny Foster"}, {"credit\_id": "549599cbc3a3686ae1004613", "department": "Crew", "gender": 0, "id": 1401791, "job": "Visual Effects Editor", "name": "Christopher Marino"}, {"credit\_id": "549599f2c3a3686ae100461e", "department": "Crew", "gender": 0, "id": 1401792, "job": "Visual Effects Editor", "name": "Jim Milton"}, {"credit\_id": "54959a51c3a3686af3003eb5", "department": "Visual Effects", "gender": 0, "id": 1401793, "job": "Visual Effects Producer", "name": "Cyndi Ochs"}, {"credit\_id": "54959a7cc3a36811530026f4", "department": "Crew", "gender": 0, "id": 1401794, "job": "Visual Effects Editor", "name": "Lucas Putnam"}, {"credit\_id": "54959b91c3a3680ff5002cb4", "department": "Visual Effects", "gender": 0, "id": 1401795, "job": "Visual Effects Supervisor", "name": "Anthony \\'Max\' Ivins"}, {"credit\_id": "54959bb69251412c4e002a5f", "department": "Visual Effects", "gender": 0, "id": 1401796, "job": "Visual Effects Supervisor", "name": "John Knoll"}, {"credit\_id": "54959cbbc3a3686ae3004391", "department": "Visual Effects", "gender": 2, "id": 1401799, "job": "Visual Effects Supervisor", "name": "Eric Saindon"}, {"credit\_id": "54959d06c3a3686ae90043f6", "department": "Visual Effects", "gender": 0, "id": 1401800, "job": "Visual Effects Supervisor", "name": "Wayne Stables"}, {"credit\_id": "54959d259251416e1e003f11", "department": "Visual Effects", "gender": 0, "id": 1401801, "job": "Visual Effects Supervisor", "name": "David Stinnett"}, {"credit\_id": "54959db49251413af7002975", "department": "Visual Effects", "gender": 0, "id": 1401803, "job": "Visual Effects Supervisor", "name": "Guy Williams"}, {"credit\_id": "54959de4c3a3681153002750", "department": "Crew", "gender": 0, "id": 1401804, "job": "Stunt Coordinator", "name": "Stuart Thorp"}, {"credit\_id": "54959ef2c3a3680fc60027f2", "department": "Lighting", "gender": 0, "id": 1401805, "job": "Best Boy Electric", "name": "Giles Coburn"}, {"credit\_id": "54959f07c3a3680fc60027f9", "department": "Camera", "gender": 2, "id": 1401806, "job": "Still Photographer", "name": "Mark Fellman"}, {"credit\_id": "54959f47c3a3681153002774", "department": "Lighting", "gender": 0, "id": 1401807, "job": "Lighting Technician", "name": "Scott Sprague"}, {"credit\_id": "54959f8cc3a36831b8001df2", "department": "Visual Effects", "gender": 0, "id": 1401808, "job": "Animation Director", "name": "Jeremy Holloman"}, {"credit\_id": "54959fa0c3a36831b8001fdb", "department": "Visual Effects", "gender": 0, "id": 1401809, "job": "Animation Director", "name": "Orlando Meunier"}, {"credit\_id": "54959fb6c3a3686af3003f54", "department": "Visual Effects", "gender": 0, "id": 1401810, "job": "Animation Director", "name": "Taisuke Tanimura"}, {"credit\_id": "54959fd2c3a36831b8001e02", "department": "Costume & Make-Up", "gender": 0, "id": 1401812, "job": "Set Costumer", "name": "Lilia Mishel Acevedo"}, {"credit\_id": "54959ff9c3a3686ae300440c", "department": "Costume & Make-Up", "gender": 0, "id": 1401814, "job": "Set Costumer", "name": "Alejandro M. Hernandez"}, {"credit\_id": "5495a0ddc3a3686ae10046fe", "department": "Editing", "gender": 0, "id": 1401815, "job": "Digital Intermediate", "name": "Marvin Hall"}, {"credit\_id": "5495a1f7c3a3686ae3004443", "department": "Production", "gender": 0, "id": 1401816, "job": "Publicist", "name": "Judy Allay"}, {"credit\_id": "5592b29fc3a36869d10002f", "department": "Crew", "gender": 0, "id": 1418381, "job": "CG Supervisor", "name": "Mike Perry"}, {"credit\_id": "5592b23a9251415df8001081", "department": "Crew", "gender": 0, "id": 1426854, "job": "CG Supervisor", "name": "Andrew Morley"}, {"credit\_id": "55491e1192514104c40002d8", "department": "Art", "gender": 0, "id": 1438901, "job": "Conceptual Design", "name": "Seth Engstrom"}, {"credit\_id": "5525d5809251417276002b06", "department": "Crew", "gender": 0, "id": 1447362, "job": "Visual Effects Art Director", "name": "Eric Oliver"}, {"credit\_id": "554427ca925141586500312a", "department": "Visual Effects", "gender": 0, "id": 1447503, "job": "Modeling", "name": "Matsune Suzuki"}, {"credit\_id": "551906889251415aab001c88", "department": "Art", "gender": 0, "id": 1447524, "job": "Art Department Manager", "name": "Paul Tobin"}, {"credit\_id": "5592af8492514152cc0010de", "department": "Costume & Make-Up", "gender": 0, "id": 1452643, "job": "Hairstylist", "name": "Roxane Griffin"}, {"credit\_id": "553d3c109251415852001318", "department": "Lighting", "gender": 0, "id": 1453938, "job": "Lighting Artist", "name": "Arun Ram-Mohan"}, {"credit\_id": "5592af4692514152d5001355", "department": "Costume & Make-Up", "gender": 0,

```
"id": 1457305, "job": "Makeup Artist", "name": "Georgia Lockhart-Adams"}, {"credit_id": "5592b2eac3a36877470012a5", "department": "Crew", "gender": 0, "id": 1466035, "job": "CG Supervisor", "name": "Thrain Shadbolt"}, {"credit_id": "5592b032c3a36877450015f1", "department": "Crew", "gender": 0, "id": 1483220, "job": "CG Supervisor", "name": "Brad Alexander"}, {"credit_id": "5592b05592514152d80012f6", "department": "Crew", "gender": 0, "id": 1483221, "job": "CG Supervisor", "name": "Shadi Almassizadeh"}, {"credit_id": "5592b090c3a36877570010b5", "department": "Crew", "gender": 0, "id": 1483222, "job": "CG Supervisor", "name": "Simon Clutterbuck"}, {"credit_id": "5592b0dbc3a368774b00112c", "department": "Crew", "gender": 0, "id": 1483223, "job": "CG Supervisor", "name": "Graeme Demmocks"}, {"credit_id": "5592b0fe92514152db0010c1", "department": "Crew", "gender": 0, "id": 1483224, "job": "CG Supervisor", "name": "Adrián Fernandes"}, {"credit_id": "5592b11f9251415df8001059", "department": "Crew", "gender": 0, "id": 1483225, "job": "CG Supervisor", "name": "Mitch Gates"}, {"credit_id": "5592b15dc3a3687745001645", "department": "Crew", "gender": 0, "id": 1483226, "job": "CG Supervisor", "name": "Jerry Kung"}, {"credit_id": "5592b18e925141645a0004ae", "department": "Crew", "gender": 0, "id": 1483227, "job": "CG Supervisor", "name": "Andy Lomas"}, {"credit_id": "5592b1bfc3a368775d0010e7", "department": "Crew", "gender": 0, "id": 1483228, "job": "CG Supervisor", "name": "Sebastian Marino"}, {"credit_id": "5592b2049251415df8001078", "department": "Crew", "gender": 0, "id": 1483229, "job": "CG Supervisor", "name": "Matthias Menz"}, {"credit_id": "5592b27b92514152d800136a", "department": "Crew", "gender": 0, "id": 1483230, "job": "CG Supervisor", "name": "Sergei Nevshupov"}, {"credit_id": "5592b2c3c3a36869e800003c", "department": "Crew", "gender": 0, "id": 1483231, "job": "CG Supervisor", "name": "Philippe Rebours"}, {"credit_id": "5592b317c3a36877470012af", "department": "Crew", "gender": 0, "id": 1483232, "job": "CG Supervisor", "name": "Michael Takarangi"}, {"credit_id": "5592b345c3a36877470012bb", "department": "Crew", "gender": 0, "id": 1483233, "job": "CG Supervisor", "name": "David Weitzberg"}, {"credit_id": "5592b37cc3a368775100113b", "department": "Crew", "gender": 0, "id": 1483234, "job": "CG Supervisor", "name": "Ben White"}, {"credit_id": "573c8e2f9251413f5d000094", "department": "Crew", "gender": 1, "id": 1621932, "job": "Stunts", "name": "Min Windle"}]
```

```
In [38]: import ast
def converter3(obj):
    l = []
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            l.append(i['name'])
            break
    return l
```

```
In [39]: movies['crew']=movies['crew'].apply(converter3)
```

```
In [40]: movies.head()
```

Out[40]:

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

In [41]: movies['overview'][0]

Out[41]: '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]: movies['overview']=movies['overview'].apply(lambda x:x.split())

In [43]: movies.head()

Out[43]:

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

In [44]: `movies['genres']=movies['genres'].apply(lambda x:[i.replace(" ", "") for i in x])`

In [45]: `movies.head()`

Out[45]:

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

In [46]: movies['keywords']=movies['keywords'].apply(lambda x:[i.replace(" ","") for i in x])

In [47]: movies['cast']=movies['cast'].apply(lambda x:[i.replace(" ","") for i in x])

In [48]: movies['crew']=movies['crew'].apply(lambda x:[i.replace(" ","") for i in x])

In [49]: movies.head()

Out[49]:

	movie_id	title	overview	keywords	genres	cast
0	1995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[cultureclash, future, spacewar, spacecolony, ...]	[Action, Adventure, Fantasy, ScienceFiction]	[SamWorthington, ZoeSaldana, SigourneyWeaver] [JamesCa
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...]	[ocean, drugabuse, exoticisland, eastindiatriad...]	[Adventure, Fantasy, Action]	[JohnnyDepp, OrlandoBloom, KeiraKnightley] [GoreVe
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[spy, basedonnovel, secretagent, sequel, mi6, ...]	[Action, Adventure, Crime]	[DanielCraig, ChristophWaltz, LéaSeydoux] [SamN
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[dccomics, crimefighter, terrorist, secretiden...]	[Action, Crime, Drama, Thriller]	[ChristianBale, MichaelCaine, GaryOldman] [Christophe
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[basedonnovel, mars, medallion, spacetravel, p...]	[Action, Adventure, ScienceFiction]	[TaylorKitsch, LynnCollins, SamanthaMorton] [AndrewS

In [50]: movies['tags'] = movies['overview'] + movies['keywords'] + movies['genres'] + movies['

In [51]: movies.head()

Out[51]:

	movie_id	title	overview	keywords	genres	cast
0	1995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[cultureclash, future, spacewar, spacecolony, ...]	[Action, Adventure, Fantasy, ScienceFiction]	[SamWorthington, ZoeSaldana, SigourneyWeaver]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...]	[ocean, drugabuse, exoticisland, eastindiatriad...]	[Adventure, Fantasy, Action]	[JohnnyDepp, OrlandoBloom, KeiraKnightley]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[spy, basedonnovel, secretagent, sequel, mi6, ...]	[Action, Adventure, Crime]	[DanielCraig, ChristophWaltz, LéaSeydoux]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[dccomics, crimefighter, terrorist, secretiden...]	[Action, Crime, Drama, Thriller]	[ChristianBale, MichaelCaine, [Christophe GaryOldman]]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[basedonnovel, mars, medallion, spacetravel, p...]	[Action, Adventure, ScienceFiction]	[TaylorKitsch, LynnCollins, SamanthaMorton]

In [52]: new\_df = movies[['movie\_id', 'title', 'tags']]

In [53]: new\_df

Out[53]:

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

```
C:\Users\JAY ACHARYA\AppData\Local\Temp\ipykernel_20300\3060717092.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
    new_df['tags']=new_df['tags'].apply(lambda x : " ".join(x))
```

In [55]: `new_df.head()`

Out[55]:

	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 [56]: `import nltk`

```
In [57]: from nltk.stem.porter import PorterStemmer  
ps = PorterStemmer()
```

```
In [58]: def stem(text):  
    y = []  
    for i in text.split():  
        y.append(ps.stem(i))  
  
    return " ".join(y)
```

```
In [59]: new_df['tags']=new_df['tags'].apply(stem)
```

```
C:\Users\JAY ACHARYA\AppData\Local\Temp\ipykernel_20300\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  
new_df['tags']=new_df['tags'].apply(stem)
```

```
In [60]: new_df['tags'][0]
```

```
Out[60]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mis  
sion, but becom torn between follow order and protect an alien civilization. culturec  
lash futur spacewar spacecoloni societi spacetravel futurist romanc space alien tribe  
alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3d action  
adventur fantasi sciencefict samworthington zoesaldana sigourneyweav jamescameron'
```

```
In [61]: new_df['tags']=new_df['tags'].apply(lambda x: x.lower())
```

```
C:\Users\JAY ACHARYA\AppData\Local\Temp\ipykernel_20300\2543325826.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  
new_df['tags']=new_df['tags'].apply(lambda x: x.lower())
```

```
In [62]: new_df.head()
```

```
Out[62]:
```

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is disp...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha...
2	206647	Spectre	a cryptic messag from bond' past send him on...
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...

```
In [63]: from sklearn.feature_extraction.text import CountVectorizer  
cv = CountVectorizer(max_features=5000,stop_words='english')
```

```
In [64]: vector=cv.fit_transform(new_df['tags']).toarray()

In [65]: cv.get_feature_names_out()

Out[65]: array(['000', '007', '10', ..., 'zone', 'zoo', 'zooeydeschanel'],
              dtype=object)

In [66]: from sklearn.metrics.pairwise import cosine_similarity

In [67]: similarity=cosine_similarity(vector)

In [68]: def recommender(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:x[1])[1:]
    for i in movies_list:
        print(new_df.iloc[i[0]].title)

In [69]: recommender("Avatar")
Aliens vs Predator: Requiem
Aliens
Falcon Rising
Independence Day
Titan A.E.
```

```
In [ ]:
```