

```
In [63]: import numpy as np
import pandas as pd
import ast
```

```
In [64]: movies=pd.read_csv('tmdb_5000_movies.csv')
credits=pd.read_csv('tmdb_5000_credits.csv')
```

```
In [66]: movies.head(1)
```

```
Out[66]:
```

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

```
In [67]: credits.head(1)
```

```
Out[67]:
```

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

```
In [68]: movies=movies.merge(credits, on='title')
```

```
In [69]: movies.head(1)
```

```
Out[69]:
```

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

1 rows × 23 columns

```
In [70]: #INCLUDED
#genres
#id
#keywords
#title
#overview
#cast
#crew

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

In [71]: movies.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4809 entries, 0 to 4808
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   movie_id    4809 non-null    int64  
 1   title        4809 non-null    object  
 2   overview     4806 non-null    object  
 3   genres       4809 non-null    object  
 4   keywords     4809 non-null    object  
 5   cast         4809 non-null    object  
 6   crew         4809 non-null    object  
dtypes: int64(1), object(6)
memory usage: 300.6+ KB
```

```
In [72]: movies.head()
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "culture clash"}]	[{"id": 1463, "name": "culture clash"}]	[{"cast_id": 242, "character": "Jake Sully", "id": 726}...]	[{"credit_id": "52fe48009251416c750aca23", "de..."}]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "ocean"}]	[{"id": 270, "name": "ocean"}]	[{"cast_id": 4, "character": "Captain Jack Spa..."}]	[{"credit_id": "52fe4232c3a36847f800b579", "de..."}]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "spy"}]	[{"id": 470, "name": "spy"}]	[{"cast_id": 1, "character": "James Bond", "name": "cr..."}]	[{"credit_id": "54805967c3a36829b5002c41", "de..."}]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "name": "dc comics"}]	[{"id": 849, "name": "dc comics"}]	[{"cast_id": 2, "character": "Bruce Wayne / Ba...", "id": 853,...}...]	[{"credit_id": "52fe4781c3a36847f81398c3", "de..."}]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "based on novel"}]	[{"id": 818, "name": "based on novel"}]	[{"cast_id": 5, "character": "John Carter", "name": "c..."}]	[{"credit_id": "52fe479ac3a36847f813eaa3", "de..."}]

```
In [73]: movies.isnull().sum()
```

```
Out[73]: movie_id      0  
          title       0  
          overview    3  
          genres      0  
          keywords    0  
          cast        0  
          crew        0  
          dtype: int64
```

```
In [74]: movies.dropna(inplace=True)
```

```
In [75]: movies.duplicated().sum()
```

```
Out[75]: 0
```

```
In [76]: movies.iloc[0].genres
```

```
Out[76]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [77]: def convert(obj):  
    L=[]  
    for i in ast.literal_eval(obj):  
        L.append(i['name'])  
    return L
```

```
In [78]: movies['genres']=movies['genres'].apply(convert)
```

```
In [79]: movies.head()
```

	Out[79]:	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]	[{"id": 1463, "name": "culture clash"}, {"id": ...]	[{"cast_id": 242, "character": "Jake Sully", "id": ...}	[{"credit_id": "52fe48009251416c750aca23", "de...]
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_id": "52fe4232c3a36847f800b579", "de...]
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_id": "54805967c3a36829b5002c41", "de...]
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_id": "52fe4781c3a36847f81398c3", "de...]
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_id": "52fe479ac3a36847f813eaa3", "de...]

In [80]: `movies['keywords'] = movies['keywords'].apply(convert)`

In [81]: `movies.head()`

Out[81]:	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...]	[{"cast_id": 242, "character": "Jake Sully", "..."}]	[{"credit_id": "52fe48009251416c750aca23", "de..."}
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_id": "52fe4232c3a36847f800b579", "de..."}
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_id": "54805967c3a36829b5002c41", "de..."}
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_id": "52fe4781c3a36847f81398c3", "de..."}
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_id": "52fe479ac3a36847f813eaa3", "de..."}

In [82]:

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

In [83]:

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

In [84]:

```
movies.head()
```

	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, ...]	{"credit_id": "52fe48009251416c750aca23", "de...
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_id": "52fe4232c3a36847f800b579", "de...
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, R...	{"credit_id": "54805967c3a36829b5002c41", "de...
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, A...	{"credit_id": "52fe4781c3a36847f81398c3", "de...
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_id": "52fe479ac3a36847f813eaa3", "de...

```
In [86]: movies['crew'][0]
```

```
Out[86]: '[{"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 DeVetta"}, {"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": "Camerawork", "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": "54959cdfc3a3681153002729", "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": "54491967c3a3680faaa001b5e", "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": "54491b79c3a3680fb001826", "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": 140178}

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 Hollobon"}, {"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 Ally"}, {"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 [89]: def fetch_director(obj):
    L = []
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            L.append(i['name'])
            break
    return L
```

```
In [91]: movies['crew']=movies['crew'].apply(fetch_director)
```

```
In [92]: movies.head()
```

	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 Weave...	[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, R...	[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, A...]	[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 [94]: `movies['overview']=movies['overview'].apply(lambda x:x.split())`

In [95]: `movies.head()`

Out[95]:	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...]	[Sam Worthington, Zoe Saldana, Sigourney Weave...]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...]	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...]	[Johnny Depp, Orlando Bloom, Keira Knightley, ...]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...]	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...]	[Christian Bale, Michael Caine, Gary Oldman, A...]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...]	[Taylor Kitsch, Lynn Collins, Samantha Morton,...]	[Andrew Stanton]

In [97]:

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

In [98]:

```
movies.head()
```

Out[98]:	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...]	[SamWorthington, ZoeSaldana, SigourneyWeaver, ...]	[JamesCameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...]	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatriad...]	[JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste...]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...]	[DanielCraig, ChristophWaltz, LéaSeydoux, Ralp...]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...]	[ChristianBale, MichaelCaine, GaryOldman, Anne...]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...]	[TaylorKitsch, LynnCollins, SamanthaMorton, Wi...]	[AndrewStanton]

In [99]: movies['tags']=movies['overview'] + movies['genres'] + movies['keywords'] + movies['ca

In [100...]: movies.head()

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...]	[SamWorthington, ZoeSaldana, SigourneyWeaver, ...]	[JamesCameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...]	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatriad...]	[JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste...]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...]	[DanielCraig, ChristophWaltz, LéaSeydoux, Ralp...]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...]	[ChristianBale, MichaelCaine, GaryOldman, Anne...]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...]	[TaylorKitsch, LynnCollins, SamanthaMorton, Wi...]	[AndrewStanton]

In [123...]

`new_df=movies[['movie_id','title','tags']]`

In [124...]

`new_df.head()`

Out[124]:

	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 [127...]

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

```
C:\Users\Samarth\AppData\Local\Temp\ipykernel_30036\487797088.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

In [128]: `new_df.head()`

	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 Barbosa, 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 [130]: `new_df['tags']=new_df['tags'].apply(lambda x:x.lower())`

```
C:\Users\Samarth\AppData\Local\Temp\ipykernel_30036\4224080999.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    new_df['tags']=new_df['tags'].apply(lambda x:x.lower())
```

In [131]: `new_df.head()`

	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 barbosa, 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 [142]: `!pip install nltk`

```
Requirement already satisfied: nltk in c:\users\samarth\anaconda3\lib\site-packages
(3.7)
Requirement already satisfied: joblib in c:\users\samarth\anaconda3\lib\site-packages
(from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\samarth\anaconda3\lib\site-
-packages (from nltk) (2022.3.15)
Requirement already satisfied: tqdm in c:\users\samarth\anaconda3\lib\site-packages
(from nltk) (4.64.0)
Requirement already satisfied: click in c:\users\samarth\anaconda3\lib\site-packages
(from nltk) (8.0.4)
Requirement already satisfied: colorama in c:\users\samarth\anaconda3\lib\site-packages
(from click->nltk) (0.4.4)
```

In [143...]: `import nltk`

In [144...]: `from nltk.stem.porter import PorterStemmer
ps=PorterStemmer()`

In [145...]: `def stem(text):
 y=[]
 for i in text.split():
 y.append(ps.stem(i))

 return " ".join(y)`

In [147...]: `new_df['tags']=new_df['tags'].apply(stem)`

```
C:\Users\Samarth\AppData\Local\Temp\ipykernel_30036\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/us  
er_guide/indexing.html#returning-a-view-versus-a-copy  
    new_df['tags']=new_df['tags'].apply(stem)
```

In [148...]: `#using vectorisation
#Bag of words method
#dimensional space of 5000 words`

In [149...]: `from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000,stop_words='english')`

In [150...]: `vectors=cv.fit_transform(new_df['tags']).toarray()`

In [151...]: `vectors`

Out[151]: `array([[0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 ...,
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0]], dtype=int64)`

In [152...]: `vectors[0]`

```
Out[152]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
In [153... cv.get_feature_names()
```

```
C:\Users\Samarth\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.  
warnings.warn(msg, category=FutureWarning)
```

```
Out[153]: ['000',
 '007',
 '10',
 '100',
 '11',
 '12',
 '13',
 '14',
 '15',
 '16',
 '17',
 '18',
 '18th',
 '19',
 '1930',
 '1940',
 '1950',
 '1960',
 '1960s',
 '1970',
 '1970s',
 '1980',
 '1990',
 '19th',
 '19thcenturi',
 '20',
 '20th',
 '24',
 '25',
 '30',
 '3d',
 '40',
 '50',
 '60',
 '70',
 'aaron',
 'aaroneckhart',
 'aarontaylor',
 'aasifmandvi',
 'abandon',
 'abduct',
 'abigailbreslin',
 'abil',
 'abl',
 'abov',
 'abus',
 'academi',
 'accept',
 'access',
 'accid',
 'accident',
 'accompani',
 'accomplish',
 'account',
 'accus',
 'ace',
 'achiev',
 'act',
 'action',
 'activ',
```

'activist',
'actor',
'actress',
'actual',
'adam',
'adambrodi',
'adamgoldberg',
'adamlefevr',
'adamsandl',
'adamscott',
'adamshankman',
'adapt',
'add',
'addict',
'adewaleakinnuoye',
'adjust',
'admir',
'admit',
'adolesc',
'adopt',
'ador',
'adrianmartinez',
'adrienbrodi',
'adult',
'adulteri',
'adulthood',
'advanc',
'adventur',
'adventure',
'advertis',
'advic',
'advis',
'affair',
'affect',
'afghanistan',
'africa',
'african',
'aftercreditssting',
'afterlif',
'aftermath',
'ag',
'agbaj',
'age',
'agediffer',
'agency',
'agenda',
'agent',
'aggress',
'ago',
'agre',
'ahead',
'aid',
'aidanquinn',
'ail',
'aim',
'air',
'airplan',
'airport',
'al',
'alanalda',

'alanarkin',
'alancum',
'alanrickman',
'alantudyk',
'alaska',
'albert',
'albertbrook',
'albertfinney',
'alcohol',
'alecbaldwin',
'alessandronivola',
'alex',
'alexandersiddig',
'alexapenavega',
'alexborstein',
'alexisbledel',
'alfredhitchcock',
'alfredmolina',
'alfrewoodard',
'ali',
'alic',
'alice',
'alicedrummond',
'aliciasilverston',
'aliciawitt',
'alien',
'alieninvas',
'alienlife',
'alilart',
'aliv',
'alive',
'allancordun',
'allen',
'allencovert',
'alleong',
'alli',
'allianc',
'allisonjanney',
'allow',
'alon',
'alongsid',
'alpacino',
'already',
'alter',
'altern',
'alway',
'alzheimer',
'amanda',
'amandapeet',
'amandaseyfri',
'amateur',
'amaz',
'ambassador',
'amberheard',
'ambit',
'ambiti',
'ambul',
'ambush',
'america',
'american',

'americanfootbal',
'amid',
'amidst',
'amnesia',
'amp',
'amus',
'amusementpark',
'amyadam',
'amypoehl',
'amyryan',
'amysedari',
'amysmart',
'anagastey',
'analyst',
'anarchiccomedi',
'ancient',
'anderson',
'andi',
'andiemacdowell',
'andreamartin',
'andrebraugh',
'andrew',
'andrewdali',
'android',
'andydick',
'andygarcía',
'andyricht',
'andysamberg',
'andyserki',
'angel',
'angelabassett',
'angeles',
'angelinajoli',
'angelo',
'anger',
'angle',
'angri',
'angusmacinn',
'ani',
'anim',
'animalattack',
'animalhorror',
'animals',
'anjelicahuston',
'ann',
'anna',
'annafari',
'annakendrick',
'annalevin',
'annaprun',
'annasophiarobb',
'anndowd',
'anne',
'annebancroft',
'annefletch',
'annehathaway',
'annehech',
'annemoss',
'annetteben',
'anni',

'annikaperga',
'announc',
'annual',
'anonym',
'anoth',
'answer',
'ant',
'antholog',
'anthoni',
'anthonyanderson',
'anthonyhopkin',
'anthonymacki',
'anthropomorph',
'anti',
'antihero',
'antonicon',
'antonibandera',
'antonyelchin',
'anyon',
'anyth',
'apart',
'apartheid',
'apartment',
'ape',
'apocalyps',
'apocalypt',
'appar',
'appear',
'appoint',
'appreci',
'apprentic',
'approach',
'april',
'arab',
'archaeologist',
'architect',
'arci',
'arctic',
'area',
'arena',
'argument',
'arielwint',
'arigraynor',
'aris',
'aristocrat',
'arm',
'armi',
'arminmueller',
'armor',
'armsdeal',
'armstrong',
'army',
'arnold',
'arnoldschwarzenegg',
'arrang',
'arrest',
'arriv',
'arrog',
'art',
'anthur',

'arthurj',
'arthurtovvey',
'artificialintellig',
'artist',
'ash',
'ashley',
'ashleyjudd',
'ashtonkutch',
'asia',
'ask',
'aspir',
'assassin',
'assault',
'assign',
'assist',
'assistant',
'associ',
'assum',
'asteroid',
'astronaut',
'asylum',
'athlet',
'atomicbomb',
'attack',
'attempt',
'attend',
'attent',
'attitud',
'attorney',
'attract',
'audienc',
'audit',
'august',
'aunjanueelli',
'aunt',
'austin',
'australia',
'australian',
'author',
'autism',
'auto',
'avaacr',
'aveng',
'avoid',
'await',
'awaken',
'award',
'away',
'awkward',
'awri',
'awry',
'azizansari',
'babe',
'babbi',
'baby',
'bachelor',
'backdrop',
'background',
'bad',
'bag',

'bahama',
'bail',
'bailey',
'balanc',
'ball',
'ballet',
'band',
'bandit',
'banish',
'bank',
'banker',
'bankrobberi',
'baptist',
'bar',
'barbarahershay',
'barbrastreisand',
'bare',
'bargain',
'barn',
'barney',
'baron',
'barri',
'barrylevinson',
'barrypepp',
'barryshabakahenley',
'bas',
'base',
'basebal',
'basedoncomicbook',
'basedongraphicnovel',
'basedonnovel',
'basedonplay',
'basedonstagemus',
'basedontrueev',
'basedontruestori',
'basedontvseri',
'basedonvideogam',
'basedonyoungadultnovel',
'basement',
'basketbal',
'batman',
'battl',
'battle',
'battlefield',
'bay',
'bdwong',
'beach',
'bear',
'beast',
'beat',
'beaubridg',
'beauti',
'beautiful',
'beautifulwoman',
'beauty',
'becam',
'becaus',
'beccasweitz',
'becom',
'becominganadult',

'bed',
'bedroom',
'beer',
'befor',
'befriend',
'began',
'begin',
'behavior',
'belief',
'believ',
'bell',
'bella',
'belong',
'belov',
'ben',
'benaffleck',
'benchaplin',
'bend',
'benedictcumberbatch',
'benedictwong',
'benfost',
'bengazzara',
'beniciodeltoro',
'benjamin',
'benjaminbratt',
'benkingsley',
'benmendelsohn',
'bennett',
'benstil',
'bent',
'benwhishaw',
'berlin',
'bernardhill',
'bernardle',
'berniemac',
'best',
'bestfriend',
'bet',
'beth',
'bethgrant',
'betray',
'bettemidl',
'better',
'betti',
'beverli',
'beverlyd',
'bibl',
'big',
'bigger',
'biggest',
'biker',
'bikini',
'billcamp',
'billcobb',
'billduk',
'billhad',
'billi',
'billionair',
'billmoseley',
'billmurray',

'billnighi',
'billnunn',
'billpaxton',
'billpullman',
'billybobthornton',
'billyburk',
'billyconnolli',
'billycrudup',
'billycryst',
'billydeewilliam',
'billyslaught',
'billywest',
'billyzan',
'biographi',
'bird',
'birth',
'birthday',
'bisexu',
'bishop',
'bit',
'bite',
'bitter',
'bizarr',
'black',
'blackmag',
'blackmail',
'blackpeopl',
'blade',
'blakeclark',
'blame',
'blend',
'blind',
'bliss',
'block',
'blond',
'blood',
'bloodi',
'bloodsplatt',
'bloodthirsti',
'blow',
'blue',
'blythedann',
'board',
'boardingschool',
'boat',
'bob',
'bobbabalaban',
'bobbergen',
'bobbi',
'bobbycannaval',
'bobbyfarrelli',
'bobbymoynihan',
'bobclendenin',
'bobgunton',
'bobhoskin',
'bobodenkirk',
'bobstephenson',
'bodi',
'body',
'bodyguard',

'bokeemwoodbin',
'bold',
'bollywood',
'bomb',
'bond',
'bone',
'bonniehunt',
'book',
'border',
'bore',
'boredom',
'born',
'boss',
'boston',
'bound',
'bounti',
'bountyhunt',
'box',
'boxer',
'boy',
'boydbank',
'boydholbrook',
'boyfriend',
'boys',
'braddourif',
'bradgarrett',
'bradleycoop',
'bradleywhitford',
'bradpitt',
'bradwilliamhenk',
'brain',
'brand',
'brave',
'braveri',
'brazil',
'brazilian',
'break',
'breckinmey',
'brendanfras',
'brendangleeson',
'brendansextoniii',
'brentbrisco',
'brentspin',
'brettcullen',
'brettratn',
'brian',
'briancox',
'briandenneh',
'briandepalma',
'briandoyle',
'brianj',
'briano',
'bride',
'bridg',
'bridgettewilson',
'brief',
'brielarson',
'brien',
'brilliant',
'bring',

'brink',
'britain',
'british',
'britishsecret servic',
'brittanymurphi',
'broadway',
'broke',
'broken',
'brook',
'brooklyn',
'brothel',
'brother',
'brotherbrotherrelationship',
'brothers',
'brothersisterrelationship',
'brought',
'brown',
'bruce',
'brucecampbel',
'brucedern',
'brucegreenwood',
'brucemcgil',
'brucesp',
'brucewilli',
'brutal',
'bryan',
'bryancranston',
'brysing',
'brycedallashoward',
'bu',
'buck',
'bud',
'buddi',
'buddycomedi',
'budget',
'build',
'building',
'built',
'bullet',
'bulli',
'bumbl',
'burglar',
'buri',
'burn',
'burtreynold',
'bush',
'busi',
'business',
'businessman',
'bust',
'butler',
'buy',
'byrn',
'cabin',
'cal',
'california',
'callumkeithrenni',
'calvin',
'came',
'camera',

'cameron',
'cameronbright',
'camerondiaz',
'camgigandet',
'camp',
'campaign',
'campbel',
'campu',
'canada',
'canadian',
'cancer',
'candid',
'cannib',
'capabl',
'capit',
'captain',
'captiv',
'captur',
'car',
'caraccid',
'caraseymour',
'carchas',
'carcrash',
'card',
'care',
'career',
'careymulligan',
'caribbean',
'carjourney',
'carl',
'carlagallo',
'carlagugino',
'carlo',
'carlosalazraqui',
'carmen',
'carmenelectra',
'carol',
'carolineaaron',
'carolinegoodal',
'carolkan',
'carrac',
'carri',
'carrie',
'carriefish',
'carter',
'cary',
'caryelw',
'case',
'caseyaffleck',
'cash',
'casino',
'cast',
'castl',
'cat',
'cataclysm',
'catastroph',
'catch',
'cateblanchett',
'catherinekeen',
'catherineo',

'catherinezeta',
'cathey',
'cathol',
'catholic',
'caught',
'caus',
'cave',
'cedrictheentertain',
'celebr',
'celiaimri',
'celiaweston',
'cellphon',
'cemeteri',
'center',
'central',
'centuri',
'century',
'certain',
'chadmichaelmurray',
'chain',
'chainsaw',
'challeng',
'champion',
'championship',
'chanc',
'chance',
'chancekelli',
'chang',
'channingtatum',
'chao',
'chaos',
'chapter',
'charact',
'character',
'characters',
'charg',
'charismat',
'charl',
'charlesd',
'charlesdurn',
'charlesnapi',
'charless',
'charli',
'charlie',
'charliesheen',
'charlizetheron',
'charlotterampl',
'charltonheston',
'charm',
'chase',
'chazzpalminteri',
'cheat',
'cheechmarin',
'cheerlead',
'chef',
'chelcieross',
'chemic',
'cher',
'chevychas',
'chicago',

'chicken',
'chief',
'child',
'childabus',
'childhood',
'children',
'chill',
'chimcbrid',
'china',
'chines',
'chip',
'chiwetelejiofor',
'chloëgracemoretz',
'chloësevigni',
'choic',
'choos',
'chosen',
'chri',
'chrisbrown',
'chriscolumbu',
'chriscoop',
'chriselli',
'chrisevan',
'chrishemsworth',
'chrisklein',
'chrismessina',
'chrismil',
'chriso',
'chrisparnel',
'chrispenn',
'chrispin',
'chrispratt',
'chrisrock',
'christ',
'christian',
'christianbal',
'christianslat',
'christinaapple',
'christinaricci',
'christinebaranski',
'christma',
'christmas',
'christmasparti',
'christoph',
'christopherguest',
'christopherlambert',
'christopherle',
'christopherlloyd',
'christophermcdonald',
'christophermeloni',
'christophermintz',
'christopherplumm',
'christopherwalken',
'christophwaltz',
'chronicl',
'church',
'cia',
'ciaránhind',
'cillianmurphi',
'cinema',

'circl',
'circu',
'circumst',
'citi',
'citizen',
'city',
'civil',
'civilian',
'civilwar',
'claim',
'clair',
'clairedan',
'claireforlani',
'clan',
'clancybrown',
'clark',
'clarkgregg',
'clash',
'class',
'classic',
'classmat',
'classroom',
'claudevandamm',
'cleaduval',
'clean',
'clear',
'clerk',
'client',
'cliffcurti',
'cliftoncollinsjr',
'climat',
'climb',
'clinteastwood',
'clinthoward',
'cliveowen',
'clock',
'clone',
'clorisleachman',
'close',
'closer',
'clown',
'club',
'clue',
'coach',
'coast',
'cocain',
'code',
'codycameron',
'coffin',
'cohen',
'col',
'cold',
'coldwar',
'cole',
'colehaus',
'colinfarrel',
'colinfirth',
'colinhank',
'colinsalmon',
'collaps',

'colleagu',
'collect',
'collector',
'colleencamp',
'colleg',
'college',
'collid',
'collin',
'collis',
'colmfeor',
'colmmeaney',
'colonel',
'coloni',
'color',
'coma',
'combat',
'combin',
'come',
'comeback',
'comedi',
'comedian',
'comedy',
'comfort',
'comic',
'coming',
'comingofag',
'comingout',
'command',
'commando',
'commerci',
'commit',
'common',
'commun',
'compani',
'companion',
'company',
'compet',
'competit',
'competition',
'complet',
'complex',
'complic',
'compos',
'compton',
'comput',
'concern',
'concert',
'condit',
'confess',
'confid',
'conflict',
'confront',
'confus',
'congress',
'conman',
'connect',
'connel',
'connienielsen',
'connieray',
'connor',

```
'conquer',
'conradvernon',
'consequ',
'consequences',
'consid',
'conspir',
'conspiraci',
'constantli',
'construct',
'consum',
'contact',
'contain',
'contemporari',
'contend',
'contest',
'continu',
'contract',
'control',
'controversi',
'convent',
'convict',
'convinc',
'cook',
'cool',
'cooper',
'cop',
'cope',
'core',
'coreyburton',
'coreystol',
'corner',
'corpor',
'corps',
'corrupt',
'cost',
'costum',
'couldn',
'count',
'countri',
'country',
...]
```

```
In [154...]: #calculating the distance between the movies
#using cosine distance
#greater the angle, more dis-similar the movie
#distance is inversely proportional to similarity
```

```
In [155...]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [157...]: similarity=cosine_similarity(vectors)
```

```
In [166...]: sorted(list(enumerate(similarity[0])), reverse=True, key=lambda x:x[1])[1:6]
```

```
Out[166]: [(1216, 0.27028123880866767),
(582, 0.23162743094465488),
(2333, 0.22996655275195),
(3730, 0.22498852128662872),
(507, 0.22438727760202976)]
```

In [174...]

```
def recommend (movie):
    movie_index=new_df[new_df['title']==movie].index[0]
    distances=similarity[movie_index]
    movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])[1:6]

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

In [176...]

```
recommend('Batman Begins')
```

The Dark Knight
Batman
The Dark Knight Rises
Batman
Rockaway

In [173...]

```
new_df.iloc[1216].title
```

Out[173]:

```
'Aliens vs Predator: Requiem'
```

In [177...]

```
import pickle
```

In [178...]

```
pickle.dump(new_df,open('movies.pkl','wb'))
```

In [179...]

```
new_df['title'].values
```

Out[179]:

```
array(['Avatar', "Pirates of the Caribbean: At World's End", 'Spectre',
       ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
       'My Date with Drew'], dtype=object)
```

In [181...]

```
pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

In [182...]

```
pickle.dump(similarity,open('similarity.pkl','wb'))
```

In []: