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

```
In [2]: movies = pd.read_csv('tmdb_5000_movies.csv')
        credits = pd.read_csv('tmdb_5000_credits.csv')
```

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

Out[3]:

	budget	genres	homepage	id	keywords	original_l
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	

In [4]: `movies.head(1)`

Out[4]:

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

◀ ████████████ ▶

In [5]: `credits.head(1)`

Out[5]:

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

In [6]: `movies.shape`

Out[6]: `(4803, 20)`

In [7]: `credits.shape`

Out[7]: `(4803, 4)`

In [8]: `movies = movies.merge(credits, on='title')`

In [9]: `movies.shape`

Out[9]: `(4809, 23)`

```
In [10]: movies.head()
```

Out[10]:

	budget	genres	homepage	id	keywords	original_la
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	

5 rows × 23 columns

```
In [11]: movies.head(1)
```

Out[11]:

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

1 rows × 23 columns

```
In [12]: #genres
         #id
         #keywords
         #title
         #overview
         #cast
         #crew


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

```
In [13]: movies.head()
```

Out[13]:

	movie_id	title	overview	genres	keywords	cast	cr
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_ "52fe48009251416c750aca2 "d
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_ "52fe4232c3a36847f800b57 "d
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_ "54805967c3a36829b5002c4 "d
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_ "52fe4781c3a36847f81398c "d
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_ "52fe479ac3a36847f813eaa "d

```
In [14]:   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 [15]:   movies.isnull().sum()
```

```
Out[15]:   movie_id    0
           title       0
           overview    3
           genres      0
           keywords    0
           cast        0
           crew        0
           dtype: int64
```

```
In [16]:   movies.dropna(inplace=True)
```

```
In [17]:   movies.duplicated().sum()
```

```
Out[17]:   0
```

```
In [18]:   movies.iloc[0].genres
```

```
Out[18]:   '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14,
           "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [19]:   #'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "n
           {"id": 878, "name": "Science Fiction"}]'
           #[action,adventure,fantasy,scifi]
```

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

```
In [21]:  movies['genres']= movies['genres'].apply(convert)
```

```
In [22]:  movies['genres']
```

```
Out[22]:  0          [Action, Adventure, Fantasy, Science Fiction]
          1                         [Adventure, Fantasy, Action]
          2                            [Action, Adventure, Crime]
          3                       [Action, Crime, Drama, Thriller]
          4                 [Action, Adventure, Science Fiction]
                                          ...
          4804                        [Action, Crime, Thriller]
          4805                              [Comedy, Romance]
          4806            [Comedy, Drama, Romance, TV Movie]
          4807                                             []
          4808                                  [Documentary]
          Name: genres, Length: 4806, dtype: object
```

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

Out[23]:

	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", "...	[{"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 [24]: movies['keywords']= movies['keywords'].apply(convert)
```

```
In [25]: movies.head()
```

Out[25]:

	movie_id	title	overview	genres	keywords	cast	c
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit, "52fe48009251416c750aca ""
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, "52fe4232c3a36847f800b5 ""
2	206647	Spectre	A cryptic message from Bond's	[Action, Adventure	[spy, based on novel, secret	[{"cast_id": 1, "character":	[{"credit, "54805967c3a36829b5002c

```
In [26]: movies['cast']
```

```
Out[26]: 0        [{"cast_id": 242, "character": "Jake Sully", "...
         1        [{"cast_id": 4, "character": "Captain Jack Spa...
         2        [{"cast_id": 1, "character": "James Bond", "cr...
         3        [{"cast_id": 2, "character": "Bruce Wayne / Ba...
         4        [{"cast_id": 5, "character": "John Carter", "c...
                                        ...
         4804     [{"cast_id": 1, "character": "El Mariachi", "c...
         4805     [{"cast_id": 1, "character": "Buzzy", "credit_...
         4806     [{"cast_id": 8, "character": "Oliver O\u2019To...
         4807     [{"cast_id": 3, "character": "Sam", "credit_id...
         4808     [{"cast_id": 3, "character": "Herself", "credi...
         Name: cast, Length: 4806, dtype: object
```

```
In [27]: import ast
         def convert3(obj):
             L=[]
             counter = 0
             for i in ast.literal_eval(obj):
                 if counter!=3:
                     L.append(i['name'])
                     counter = counter+1
                 else:
                     break
             return L
```

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

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

Out[29]:

	movie_id	title	overview	genres	keywords	cast	cre
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_i "52fe48009251416c750aca2: "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_i "52fe4232c3a36847f800b57: "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]	[{"credit_i "54805967c3a36829b5002c4 "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]	[{"credit_i "52fe4781c3a36847f81398c: "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_i "52fe479ac3a36847f813eaa: "de

```
In [30]: movies['crew'][0]
```
der": 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"}, {"credi
t_id": "52fe48009251416c750aca2f", "department": "Production", "gender": 2,
"id": 8529, "job": "Producer", "name": "Jon Landau"}, {"credit_id": "539c49
37c3a36810ba002194", "department": "Art", "gender": 0, "id": 9618, "job":
"Art Direction", "name": "Sean Haworth"}, {"credit_id": "539c49b6c3a36810c1
0020e6", "department": "Art", "gender": 1, "id": 12653, "job": "Set Decorat
ion", "name": "Kim Sinclair"}, {"credit_id": "570b6f2f9251413a0e00020d", "d
epartment": "Art", "gender": 1, "id": 12653, "job": "Supervising Art Direct
or", "name": "Kim Sinclair"}, {"credit_id": "54491a6c0e0a26748c001b19", "de
partment": "Art", "gender": 2, "id": 14350, "job": "Set Designer", "name":
"Richard F. Mays"}, {"credit_id": "56928cf4c3a3684cff0025c4", "department":
"Production", "gender": 1, "id": 20294, "job": "Executive Producer", "nam
e": "Laeta Kalogridis"}, {"credit_id": "52fe48009251416c750aca51", "departm
ent": "Costume & Make-Up", "gender": 0, "id": 17675, "job": "Costume Desig
n", "name": "Mayes C. Rubeo"}, {"credit_id": "52fe48009251416c750aca11", "d
epartment": "Camera", "gender": 2, "id": 18265, "job": "Director of Photogr
aphy", "name": "Mauro Fiore"}, {"credit_id": "5449194d0e0a26748f001b39", "d
epartment": "Art", "gender": 0, "id": 42281, "job": "Set Designer", "name":

```python
In [31]: import ast
         def fetch_director(obj):
             L=[]
             for i in ast.literal_eval(obj):
                 if i['job']=='Director':
                     L.append(i['name'])
                     break
             return L
```

```python
In [32]: movies['crew']= movies['crew'].apply(fetch_director)
```

```
In [33]:  movies.head()
```

Out[33]:

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

```
In [34]:  movies['overview'][0]
```

Out[34]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora o
n a unique mission, but becomes torn between following orders and protecting
an alien civilization.'

```
In [35]:  movies['overview']= movies['overview'].apply(lambda x:x.split())
```

```
In [36]: movies.head()
```

Out[36]:

	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 Weaver]	[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]	[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]	[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 [37]: movies['genres'] = movies['genres'].apply(lambda x:[i.replace(" ", "") for i in
         movies['keywords'] = movies['keywords'].apply(lambda x:[i.replace(" ", "") for
         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 [38]: movies.head()
```

Out[38]:

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

```
In [39]: movies['tags'] = movies['overview']+movies['genres']+
         movies['keywords']+movies['cast']+movies['crew']
```

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

Out[40]:

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

```
In [41]: movies['tags'][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.',
          'Action',
          'Adventure',
          'Fantasy',
          'ScienceFiction',
          'cultureclash',
          'future',
          'spacewar',
          'spacecolony',
          'society',
          'spacetravel',
          'futuristic',
          'romance',
          'space',
          'alien',
          'tribe',
          'alienplanet',
          'cgi',
          'marine',
          'soldier',
          'battle',
          'loveaffair',
          'antiwar',
          'powerrelations',
          'mindandsoul',
          '3d',
          'SamWorthington',
          'ZoeSaldana',
```

```
                   'SigourneyWeaver',
                   'JamesCameron']
```

In [42]:
```
new_df = movies[['movie_id', 'title', 'tags']]
```

In [43]:
```
new_df
```

Out[43]:

	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 [44]:
```
new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x))
```

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_10512\1824047427.py:1: SettingWi
thCopyWarning:
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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x))
```

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

Out[45]:

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

```
C:\Users\lenovo\AppData\Local\Temp\ipykernel_10512\3214958533.py:1: SettingWi
thCopyWarning:
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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

In [47]: `new_df`

Out[47]:

	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...
...	...	...	...
4804	9367	El Mariachi	el mariachi just wants to play his guitar and ...
4805	72766	Newlyweds	a newlywed couple's honeymoon is upended by th...
4806	231617	Signed, Sealed, Delivered	"signed, sealed, delivered" introduces a dedic...
4807	126186	Shanghai Calling	when ambitious new york attorney sam is sent t...
4808	25975	My Date with Drew	ever since the second grade when he first saw ...

4806 rows × 3 columns

```
In [48]: new_df['tags'][0]

Out[48]: 'in the 22nd century, a paraplegic marine is dispatched to the moon pandora o
         n a unique mission, but becomes torn between following orders and protecting
         an alien civilization. action adventure fantasy sciencefiction cultureclash f
         uture spacewar spacecolony society spacetravel futuristic romance space alien
         tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrelations
         mindandsoul 3d samworthington zoesaldana sigourneyweaver jamescameron'

In [49]: new_df['tags'][1]

Out[49]: "captain barbossa, long believed to be dead, has come back to life and is hea
         ded to the edge of the earth with will turner and elizabeth swann. but nothin
         g is quite as it seems. adventure fantasy action ocean drugabuse exoticisland
         eastindiatradingcompany loveofone'slife traitor shipwreck strongwoman ship al
         liance calypso afterlife fighter pirate swashbuckler aftercreditsstinger john
         nydepp orlandobloom keiraknightley goreverbinski"

In [50]: from sklearn.feature_extraction.text import CountVectorizer
         cv = CountVectorizer(max_features = 5000, stop_words= 'english')

In [51]: vectors = cv.fit_transform(new_df['tags']).toarray()

In [52]: vectors[0]

Out[52]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)

In [53]: cv.get_feature_names()
```

C:\ProgramData\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_nam
es_out instead.
  warnings.warn(msg, category=FutureWarning)

```
Out[53]: ['000',
          '007',
          '10',
          '100',
          '11',
          '12',
          '13',
          '14',
          '15',
          '16',
          '17',
          '18',
          '18th',
          '19',

In [54]: import nltk
```

```
In [55]:  from nltk.stem.porter import PorterStemmer
          ps = PorterStemmer()
```

```
In [56]:  ps.stem('loving')
```

Out[56]: 'love'

```
In [57]:  def stem(text):
              y = []

              for i in text.split():
                  y.append(ps.stem(i))

              return " ".join(y)
```

```
In [58]:  new_df['tags'] = new_df['tags'].apply(stem)
```

C:\Users\lenovo\AppData\Local\Temp\ipykernel_10512\3213734980.py:1: SettingWi
thCopyWarning:
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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  new_df['tags'] = new_df['tags'].apply(stem)

```
In [59]:  new_df
```

Out[59]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond' past send him on a...
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...
...	...	...	...
4804	9367	El Mariachi	el mariachi just want to play hi guitar and ca...
4805	72766	Newlyweds	a newlyw couple' honeymoon is upend by the arr...
4806	231617	Signed, Sealed, Delivered	"signed, sealed, delivered" introduc a dedic q...
4807	126186	Shanghai Calling	when ambiti new york attorney sam is sent to s...
4808	25975	My Date with Drew	ever sinc the second grade when he first saw h...

4806 rows × 3 columns

```python
In [60]: from sklearn.metrics.pairwise import cosine_similarity
```

```python
In [61]: similarity = cosine_similarity(vectors)
```

```python
In [62]: similarity[2]
```
Out[62]: array([0.05976143, 0.0625   , 1.   , ..., 0.02635231, 0.   ,
         0.   ])

```python
In [63]: similarity[0]
```
Out[63]: array([1.   , 0.08964215, 0.05976143, ..., 0.02519763, 0.02817181,
         0.   ])

```python
In [64]: sorted(similarity[0], reverse = True)
```
Out[64]: [1.0,
 0.26089696604360174,
 0.2581988897471611,
 0.25302403842552984,
 0.25110592822973776,
 0.24944382578492943,
 0.24397501823713333,
 0.243599382882345,
 0.23904572186687872,
 0.2344036154692477,
 0.2267786838055363,
 0.2238868314198225,
 0.22230800575069135,
 0.22190115272469205,
 0.22131333406899523,
 0.2193633988428327,
 0.2173253797873328,
 0.21653278478430668,
 0.2164218276749025,

```
In [65]: list(enumerate(similarity[0]))

Out[65]: [(0, 1.0),
          (1, 0.08964214570007951),
          (2, 0.059761430466719674),
          (3, 0.03984095364447979),
          (4, 0.18257418583505536),
          (5, 0.1183452670827877),
          (6, 0.02258769757263128),
          (7, 0.1763834207376394),
          (8, 0.061721339984836754),
          (9, 0.07644707871564382),
          (10, 0.11769797726729994),
          (11, 0.081199794294115),
          (12, 0.09583148474999097),
          (13, 0.04733810683311508),
          (14, 0.11428571428571427),
          (15, 0.05096471914376255),
          (16, 0.08226127456606226),
          (17, 0.15026857675938213),
          (18, 0.1129384878631564),
```

```python
In [66]: 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 = lambd

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

```python
In [75]: recommend('Avatar')

         Titan A.E.
         Small Soldiers
         Independence Day
         Ender's Game
         Aliens vs Predator: Requiem
```

```python
In [68]: new_df.iloc[539].title

Out[68]: 'Titan A.E.'
```

```python
In [69]: sorted(list(enumerate(similarity[0])), reverse = True, key=lambda x:x[1])[1:6]

Out[69]: [(539, 0.26089696604360174),
          (1194, 0.2581988897471611),
          (507, 0.25302403842552984),
          (260, 0.25110592822973776),
          (1216, 0.24944382578492943)]
```

```python
In [70]: new_df[new_df['title']=="Batman Begins"].index[0]

Out[70]: 119
```

```
In [71]: import pickle
```

```
In [72]: pickle.dump(new_df.to_dict(),open('movie_dict.pkl', 'wb'))
```

```
In [73]: new_df
```

Out[73]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond' past send him on a...
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...
...	...	...	...
4804	9367	El Mariachi	el mariachi just want to play hi guitar and ca...
4805	72766	Newlyweds	a newlyw couple' honeymoon is upend by the arr...
4806	231617	Signed, Sealed, Delivered	"signed, sealed, delivered" introduc a dedic q...
4807	126186	Shanghai Calling	when ambiti new york attorney sam is sent to s...
4808	25975	My Date with Drew	ever sinc the second grade when he first saw h...

4806 rows × 3 columns

```
In [74]: pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]:
```

```
In [ ]:
```