Importing Libraries

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

Importing Dataset

```
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_language	original_title	overview
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	en	John Carter	John Carter is a war-weary, former military ca...

```
In [4]: movies.shape
```

Out[4]: (4803, 20)

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

Out[5]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [6]: credits.shape
```

Out[6]: (4803, 4)

Merging the dataframes on the basis of 'title'

In [7]: `movies = movies.merge(credits, on = 'title')`

In [8]: `movies.head()`

Out[8]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	en	John Carter	John Carter is a war-weary, former military ca...

5 rows × 23 columns

In [9]: `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
```

Extracting relevant columns

In [10]:
```python
#content based recommendation

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

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

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

Out[11]:

	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, "nam...	[{"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...	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	[{"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...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"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...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"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...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

Dealing with missing and duplicate data

In [12]:
```python
movies.isnull().sum()
```

Out[12]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [13]:
```python
movies.dropna(inplace = True)
```

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

Out[14]:
```
movie_id    0
title       0
overview    0
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [15]:
```python
movies.duplicated().sum()
```

Out[15]: 0

Extracting 'genres'

In [16]:
```python
movies.iloc[0].genres
```

Out[16]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'

In [17]:
```python
# Convert this
#'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "na
# To this
```

```
# ['Action', 'Adventure', 'Fantasy', 'Science Fiction']
```

```
In [18]:  import ast
```

```
In [19]:  def convert(obj):
              L = []
              for i in ast.literal_eval(obj):
                  L.append(i['name'])
              return L
```

```
In [20]:  movies['genres'] = movies['genres'].apply(convert)
```

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

Out[21]:

	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...

Extracting 'keywords'

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

```
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]	[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...

Extracting top 3 'cast' of the movie

```
In [24]:  def convert2(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 [25]:
```
movies['cast'] = movies['cast'].apply(convert2)
```

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

Out[26]:

	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]	[{"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]	[{"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...

Extracting Director name from the 'crew'

In [27]:
```
def fetch_director(obj):
    L = []
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            L.append(i['name'])
            break
    return L
```

In [28]:
```
movies['crew'] = movies['crew'].apply(fetch_director)
```

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

Out[29]:

	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]

Converting string type 'overview' into list type 'overview'

In [30]:
```
movies['overview'][0]
```

Out[30]: '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 [31]:
```
movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

In [32]:
```
movies.head()
```

	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]

Dealing with Whitespaces

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

Out[34]:

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

In [35]:
```python
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 4806 entries, 0 to 4808
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   movie_id   4806 non-null   int64
 1   title      4806 non-null   object
 2   overview   4806 non-null   object
 3   genres     4806 non-null   object
 4   keywords   4806 non-null   object
 5   cast       4806 non-null   object
 6   crew       4806 non-null   object
dtypes: int64(1), object(6)
memory usage: 429.4+ KB
```

Creating a new coloumn as 'tags'

In [36]:
```python
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
```

In [37]:
```python
movies.head()
```

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

Creating new datframe

```python
In [38]:  new_df = movies[['movie_id', 'title', 'tags']]
```

```python
In [39]:  new_df.head()
```

Out[39]:

	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...

Converting list type 'tags' into string type 'tags'

```python
In [40]:  new_df['tags'] = new_df['tags'].apply(lambda x :" ".join(x))
```

```
C:\Users\Uttam\AppData\Local\Temp\ipykernel_9988\4095421860.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#retu
rning-a-view-versus-a-copy
  new_df['tags'] = new_df['tags'].apply(lambda x :" ".join(x))
```

```python
In [41]:  new_df.head()
```

Out[41]:

	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...

```python
In [42]:  new_df["tags"][0]
```

```
Out[42]:  'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes to
          rn between following orders and protecting an alien civilization. Action Adventure Fantasy ScienceFiction cultu
          reclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi ma
          rine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver
          JamesCameron'
```

Converting 'tags' into lowercase

```
In [43]:  new_df['tags'] = new_df['tags'].apply(lambda x : x.lower())
```

C:\Users\Uttam\AppData\Local\Temp\ipykernel_9988\3679624391.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#retu
rning-a-view-versus-a-copy
  new_df['tags'] = new_df['tags'].apply(lambda x : x.lower())

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

Out[44]:

	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...

Stemming words in 'tags'

Simplifying words to their most basic form is called stemming. It reduces the words like "chocolates", "chocolatey", and "choco" to the root word, "chocolate".

```
In [45]:  import nltk
```

```
In [46]:  from nltk.stem.porter import PorterStemmer
          ps = PorterStemmer()
```

```
In [47]:  def stem(text):
              y = []

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

              return " ".join(y)
```

```
In [48]:  new_df['tags'].apply(stem)
```

```
Out[48]:  0        in the 22nd century, a parapleg marin is dispa...
          1        captain barbossa, long believ to be dead, ha c...
          2        a cryptic messag from bond' past send him on a...
          3        follow the death of district attorney harvey d...
          4        john carter is a war-weary, former militari ca...
                                         ...
          4804     el mariachi just want to play hi guitar and ca...
          4805     a newlyw couple' honeymoon is upend by the arr...
          4806     "signed, sealed, delivered" introduc a dedic q...
          4807     when ambiti new york attorney sam is sent to s...
          4808     ever sinc the second grade when he first saw h...
          Name: tags, Length: 4806, dtype: object
```

Text Vectorization of coloumn 'tags' without stop words

Stop words are a set of commonly used words in a language. Examples of stop words in English are "a," "the," "is," "are," etc.

```
In [49]:  from sklearn.feature_extraction.text import CountVectorizer
          cv = CountVectorizer(max_features = 5000, stop_words = 'english')
```

```
In [50]:  vectors = cv.fit_transform(new_df['tags']).toarray()
```

```
In [51]:  vectors
```

```
Out[51]: 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 [52]: vectors[0]
```

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

```
In [53]: a = cv.get_feature_names_out()
```

```
In [54]: for feature_name in a[10:20]:
             print(feature_name)
```

```
17
18
18th
19
1930s
1940s
1950
1950s
1960s
1970s
```

```
In [55]: print(len(a))
```

```
5000
```

Measuring Cosine similarity

It is a way to measure how similar two vectors are in an inner product space.

```
In [56]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [57]: similarity = cosine_similarity(vectors)
```

```
In [58]: similarity[0]
```

```
Out[58]: array([1.        , 0.08964215, 0.05976143, ..., 0.02519763, 0.02817181,
                0.        ])
```

Movie Recommendation Funtion

```
In [59]: 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 [60]: recommend('Avatar')
```

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

Sending list of movies to the website

```
In [61]: import pickle
```

```
In [62]: pickle.dump(new_df,open('movies.pkl', 'wb'))
```

```
In [63]: pickle.dump(similarity,open('similarity.pkl', 'wb'))
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js