In [2]:
```python
import pandas as pd
import numpy as np
```

In [4]:
```python
from sklearn.feature_extraction.text import CountVectorizer
```

In [5]:
```python
from sklearn.metrics.pairwise import cosine_similarity
```

In [6]:
```python
def get_title_from_index(index):
        return df[df.index == index]["title"].values[0]
```

In [7]:
```python
def get_index_from_title(title):
        return df[df.title == title]["index"].values[0]
```

In [11]:
```python
url="movie_dataset.csv"
```

In [12]:
```python
import os
```

In [13]:
```python
os.getcwd()
```

Out[13]:
```
'C:\\Users\\Mahima-PC\\2\\Assignment2\\Mini Project1'
```

In [15]:
```python
df = pd.read_csv(url)
print(df)
```

```
       index      budget                                 genres  \
0          0   237000000    Action Adventure Fantasy Science Fiction
1          1   300000000                   Adventure Fantasy Action
2          2   245000000                     Action Adventure Crime
3          3   250000000                 Action Crime Drama Thriller
4          4   260000000           Action Adventure Science Fiction
...      ...         ...                                        ...
4798    4798      220000                      Action Crime Thriller
4799    4799        9000                             Comedy Romance
4800    4800           0             Comedy Drama Romance TV Movie
4801    4801           0                                        NaN
4802    4802           0                                Documentary

                                             homepage      id  \
0                       http://www.avatarmovie.com/   19995
1          http://disney.go.com/disneypictures/pirates/     285
2           http://www.sonypictures.com/movies/spectre/  206647
3                   http://www.thedarkknightrises.com/   49026
4               http://movies.disney.com/john-carter   49529
...                                               ...     ...
4798                                             NaN    9367
4799                                             NaN   72766
4800  http://www.hallmarkchannel.com/signedsealeddel...  231617
4801                       http://shanghaicalling.com/  126186
4802                                             NaN   25975

                                            keywords original_language  \
0     culture clash future space war space colony so...                en
1     ocean drug abuse exotic island east india trad...                en
2            spy based on novel secret agent sequel mi6                en
3     dc comics crime fighter terrorist secret ident...                en
4     based on novel mars medallion space travel pri...                en
...                                              ...               ...
4798  united states\u2013mexico barrier legs arms pa...                es
4799                                             NaN                en
4800  date love at first sight narration investigati...                en
4801                                             NaN                en
4802          obsession camcorder crush dream girl                en

                          original_title  \
0                                  Avatar
1     Pirates of the Caribbean: At World's End
2                                 Spectre
3                    The Dark Knight Rises
4                             John Carter
...                                    ...
4798                            El Mariachi
4799                               Newlyweds
4800                  Signed, Sealed, Delivered
4801                        Shanghai Calling
4802                       My Date with Drew

                                            overview  popularity  ...  \
0     In the 22nd century, a paraplegic Marine is di...  150.437577  ...
1     Captain Barbossa, long believed to be dead, ha...  139.082615  ...
2     A cryptic message from Bond's past sends him o...  107.376788  ...
3     Following the death of District Attorney Harve...  112.312950  ...
4     John Carter is a war-weary, former military ca...   43.926995  ...
...                                              ...         ...  ...
4798  El Mariachi just wants to play his guitar and ...   14.269792  ...
```

```
4799   A newlywed couple's honeymoon is upended by th...     0.642552  ...
4800   "Signed, Sealed, Delivered" introduces a dedic...     1.444476  ...
4801   When ambitious New York attorney Sam is sent t...     0.857008  ...
4802   Ever since the second grade when he first saw ...     1.929883  ...

        runtime                           spoken_languages    status   \
0        162.0   [{"iso_639_1": "en", "name": "English"}, {"iso...  Released
1        169.0            [{"iso_639_1": "en", "name": "English"}]  Released
2        148.0   [{"iso_639_1": "fr", "name": "Fran\u00e7ais"},...  Released
3        165.0            [{"iso_639_1": "en", "name": "English"}]  Released
4        132.0            [{"iso_639_1": "en", "name": "English"}]  Released
...        ...                                                ...       ...
4798      81.0      [{"iso_639_1": "es", "name": "Espa\u00f1ol"}]  Released
4799      85.0                                                 []  Released
4800     120.0            [{"iso_639_1": "en", "name": "English"}]  Released
4801      98.0            [{"iso_639_1": "en", "name": "English"}]  Released
4802      90.0            [{"iso_639_1": "en", "name": "English"}]  Released

                                                tagline   \
0                           Enter the World of Pandora.
1          At the end of the world, the adventure begins.
2                              A Plan No One Escapes
3                                  The Legend Ends
4              Lost in our world, found in another.
...                                                 ...
4798   He didn't come looking for trouble, but troubl...
4799   A newlywed couple's honeymoon is upended by th...
4800                                               NaN
4801                         A New Yorker in Shanghai
4802                                               NaN

                                       title vote_average vote_count   \
0                                     Avatar          7.2      11800
1         Pirates of the Caribbean: At World's End     6.9       4500
2                                    Spectre          6.3       4466
3                          The Dark Knight Rises       7.6       9106
4                                John Carter          6.1       2124
...                                      ...          ...        ...
4798                               El Mariachi          6.6        238
4799                                Newlyweds          5.9          5
4800                   Signed, Sealed, Delivered       7.0          6
4801                          Shanghai Calling         5.7          7
4802                          My Date with Drew        6.3         16

                                                cast   \
0       Sam Worthington Zoe Saldana Sigourney Weaver S...
1       Johnny Depp Orlando Bloom Keira Knightley Stel...
2       Daniel Craig Christoph Waltz L\u00e9a Seydoux ...
3       Christian Bale Michael Caine Gary Oldman Anne ...
4       Taylor Kitsch Lynn Collins Samantha Morton Wil...
...                                                 ...
4798   Carlos Gallardo Jaime de Hoyos Peter Marquardt...
4799   Edward Burns Kerry Bish\u00e9 Marsha Dietlein ...
4800   Eric Mabius Kristin Booth Crystal Lowe Geoff G...
4801   Daniel Henney Eliza Coupe Bill Paxton Alan Ruc...
4802   Drew Barrymore Brian Herzlinger Corey Feldman ...

                                       crew          director
0       [{'name': 'Stephen E. Rivkin', 'gender': 0, 'd...    James Cameron
1       [{'name': 'Dariusz Wolski', 'gender': 2, 'depa...    Gore Verbinski
```

```
2      [{'name': 'Thomas Newman', 'gender': 2, 'depar...        Sam Mendes
3      [{'name': 'Hans Zimmer', 'gender': 2, 'departm...  Christopher Nolan
4      [{'name': 'Andrew Stanton', 'gender': 2, 'depa...     Andrew Stanton
...                                                  ...                ...
4798   [{'name': 'Robert Rodriguez', 'gender': 0, 'de...  Robert Rodriguez
4799   [{'name': 'Edward Burns', 'gender': 2, 'depart...       Edward Burns
4800   [{'name': 'Carla Hetland', 'gender': 0, 'depar...        Scott Smith
4801   [{'name': 'Daniel Hsia', 'gender': 2, 'departm...        Daniel Hsia
4802   [{'name': 'Clark Peterson', 'gender': 2, 'depa...   Brian Herzlinger

[4803 rows x 24 columns]
```

In [16]:
```python
features = ['keywords','cast','genres','director']
##Step 3: Create a column in DF which combines all selected features
for feature in features:
        df[feature] = df[feature].fillna('')
```

In [18]:
```python
def combine_features(row):
        try:
                return row['keywords'] +" "+row['cast']+" "+row["genres"]+" "+row["dir
        except:
                print ("Error:", row)
```

In [19]:
```python
df["combined_features"] = df.apply(combine_features,axis=1)
```

In [20]:
```python
cv = CountVectorizer()

count_matrix = cv.fit_transform(df["combined_features"])
```

In [25]:
```python
cosine_sim = cosine_similarity(count_matrix)
movie_user_likes = "Lockout"
```

In [22]:
```python
movie_index = get_index_from_title(movie_user_likes)

similar_movies =  list(enumerate(cosine_sim[movie_index]))
```

In [26]:
```python
sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)

i=0
for element in sorted_similar_movies:
                print (get_title_from_index(element[0]))
                i=i+1
                if i>50:
                        break
```

```
Avatar
Guardians of the Galaxy
Aliens
Star Wars: Clone Wars: Volume 1
Star Trek Into Darkness
Star Trek Beyond
Alien
Lockout
Jason X
The Helix... Loaded
Moonraker
Planet of the Apes
Galaxy Quest
Gravity
Alien³
Jupiter Ascending
The Wolverine
Silent Running
Zathura: A Space Adventure
Trekkies
Cargo
Wing Commander
Star Trek
Lost in Space
Babylon A.D.
The Fifth Element
Oblivion
Titan A.E.
AVP: Alien vs. Predator
The Empire Strikes Back
Dragonball Evolution
Superman Returns
Divergent
John Carter
The Black Hole
The Ice Pirates
Memoirs of an Invisible Man
Starship Troopers
The Astronaut's Wife
Machete Kills
Soldier
The Abyss
Damnation Alley
Men in Black
Space Cowboys
Space Dogs
The Time Machine
Sheena
Captain America: Civil War
Star Trek: Insurrection
Oz: The Great and Powerful
```

In [ ]: