

In [1]:

```
# Importing relevent libraries:

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

path = './movies.csv'
data = pd.read_csv(path)
data.head()
```

Out[1]:

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS	VOTES	RunTime	Gross
0	Blood Red Sky	(2021)	\nAction, Horror, Thriller	6.1	\nA woman with a mysterious illness is forced ...	\n Director:\nPeter Thorwarth\n\n \n Star...	21,062	121.0	NaN
1	Masters of the Universe: Revelation	(2021- )	\nAnimation, Action, Adventure	5.0	\nThe war for Eternia begins again in what may...	\n \n Stars:\nChris Wood, \nSara...	17,870	25.0	NaN
2	The Walking Dead	(2010-2022)	\nDrama, Horror, Thriller	8.2	\nSheriff Deputy Rick Grimes wakes up from a c...	\n \n Stars:\nAndrew Lincoln, \n...	885,805	44.0	NaN
3	Rick and Morty	(2013- )	\nAnimation, Adventure, Comedy	9.2	\nAn animated series that follows the exploits...	\n \n Stars:\nJustin Roiland, \n...	414,849	23.0	NaN
4	Army of Thieves	(2021)	\nAction, Crime, Horror	NaN	\nA prequel, set before the events of Army of ...	\n Director:\nMatthias Schweighöfer\n\n \n ...	NaN	NaN	NaN

In [2]:

```
# DATA CLEANING :-
# Dropping duplicates with respect to MOVIES column:

data = data.drop_duplicates(subset = ["MOVIES"], keep = False)
data
```

Out[2]:

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS	VOTES	RunTime	Gross
0	Blood Red Sky	(2021)	\nAction, Horror, Thriller	6.1	\nA woman with a mysterious illness is forced ...	\n Director:\nPeter Thorwarth\n\n \n Star...	21,062	121.0	NaN
1	Masters of the Universe: Revelation	(2021- )	\nAnimation, Action, Adventure	5.0	\nThe war for Eternia begins again in what may...	\n \n Stars:\nChris Wood, \nSara...	17,870	25.0	NaN
2	The Walking Dead	(2010-2022)	\nDrama, Horror, Thriller	8.2	\nSheriff Deputy Rick Grimes wakes up from a c...	\n \n Stars:\nAndrew Lincoln, \n...	885,805	44.0	NaN
3	Rick and Morty	(2013- )	\nAnimation, Adventure, Comedy	9.2	\nAn animated series that follows the exploits...	\n \n Stars:\nJustin Roiland, \n...	414,849	23.0	NaN
4	Army of Thieves	(2021)	\nAction, Crime, Horror	NaN	\nA prequel, set before the events of Army of ...	\n \n Director:\nMatthias Schweighöfer\n\n \n ...	NaN	NaN	NaN

...	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS	VOTES	RunTime	Gross
9845	Disparu à jamais	(2021)	Crime, Drama, Mystery	NaN	Add a Plot	Director: Juan Carlos Medina   Stars: S...	NaN	NaN	NaN
9901	Modern Family	(2009–2020)	Comedy, Drama, Romance	8.2	Jay must adapt to his young new wife, Gloria...	Director: Jason Winer   Stars: n...	3,404	23.0	NaN
9993	Totenfrau	(2022–)	Drama, Thriller	NaN	Add a Plot	Director: Nicolai Rohde   Stars: ...	NaN	NaN	NaN
9995	Arcane	(2021–)	Animation, Action, Adventure	NaN	Add a Plot		NaN	NaN	NaN
9996	Heart of Invictus	(2022–)	Documentary, Sport	NaN	Add a Plot	Director: Orlando von Einsiedel   Stars: n ...	NaN	NaN	NaN

6398 rows x 9 columns

In [3]:

```
# Removing '\n' from GENRE, ONE-LINE and STARS:

for column in ['GENRE', 'ONE-LINE', 'STARS']:
    data[column]=data[column].str.replace('\n', ' ')

# Replacing unwanted strings and converting datatype of VOTES:

data['ONE-LINE']=data['ONE-LINE'].str.replace('Add a Plot', 'Not Specified')
data['VOTES']=data['VOTES'].str.replace(', ', '')
data['VOTES']=pd.to_numeric(data['VOTES'])
data
```

Out[3]:

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS	VOTES	RunTime	Gross
0	Blood Red Sky	(2021)	Action, Horror, Thriller	6.1	A woman with a mysterious illness is forced i...	Director: Peter Thorwarth   Stars: P...	21062.0	121.0	NaN
1	Masters of the Universe: Revelation	(2021–)	Animation, Action, Adventure	5.0	The war for Eternia begins again in what may ...	Stars: Chris Wood, Sarah Mi...	17870.0	25.0	NaN
2	The Walking Dead	(2010–2022)	Drama, Horror, Thriller	8.2	Sheriff Deputy Rick Grimes wakes up from a co...	Stars: Andrew Lincoln, Norm...	885805.0	44.0	NaN
3	Rick and Morty	(2013–)	Animation, Adventure, Comedy	9.2	An animated series that follows the exploits ...	Stars: Justin Roiland, Chri...	414849.0	23.0	NaN
4	Army of Thieves	(2021)	Action, Crime, Horror	NaN	A prequel, set before the events of Army of t...	Director: Matthias Schweighöfer   Stars: St...	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
9845	Disparu à jamais	(2021)	Crime, Drama, Mystery	NaN	Not Specified	Director: Juan Carlos Medina   Stars: ...	NaN	NaN	NaN
9901	Modern Family	(2009–2020)	Comedy, Drama, Romance	8.2	Jay must adapt to his young new wife, Gloria ...	Director: Jason Winer   Stars: Ed O'...	3404.0	23.0	NaN
9993	Totenfrau	(2022–)	Drama, Thriller	NaN	Not Specified	Director: Nicolai Rohde   Stars: Fel...	NaN	NaN	NaN
9995	Arcane	(2021–)	Animation, Action, Adventure	NaN	Not Specified		NaN	NaN	NaN
		(2022–)	Documentary, Sport			Director: Orlando			

9996	Heart of Invictus	(2022-)	Documentary, Sport	NaN	Not Specified	von Einsiedel, St...	NaN	NaN	NaN	NaN	NaN
	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS	VOTES	RunTime	Gross		

6398 rows x 9 columns

In [4]:

```
# Seperating Directors and Stars into different columns from STARS:

def seperate_director(director):
    if 'Director' in director or 'Directors' in director:
        director = director.strip().split('|')[0]
        return director.split(":")[1].strip()
    else:
        return ''

def seperate_stars(stars):
    if 'Star' not in stars or 'Stars' not in stars:
        return ''
    else:
        return stars.split(":")[-1].strip()

data['Director'] = data['STARS'].apply(lambda d: seperate_director(d))
data['Stars'] = data['STARS'].apply(lambda s: seperate_stars(s))
data.head()
```

Out[4]:

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS	VOTES	RunTime	Gross	Director	Star
0	Blood Red Sky	(2021)	Action, Horror, Thriller	6.1	A woman with a mysterious illness is forced i...	Director: Peter Thorwarth   Stars: P...	21062.0	121.0	NaN	Peter Thorwarth	Pei Baumeister, Carl Anton Koch, Alexander .
1	Masters of the Universe: Revelation	(2021- )	Animation, Action, Adventure	5.0	The war for Eternia begins again in what may ...	Stars: Chris Wood, Sarah Mi...	17870.0	25.0	NaN		Chris Wood, Sarah Michell Gellar, Len Head.
2	The Walking Dead	(2010-2022)	Drama, Horror, Thriller	8.2	Sheriff Deputy Rick Grimes wakes up from a co...	Stars: Andrew Lincoln, Norm...	885805.0	44.0	NaN		Andrew Lincoln, Norma Reedus, Meliss McBri.
3	Rick and Morty	(2013- )	Animation, Adventure, Comedy	9.2	An animated series that follows the exploits ...	Stars: Justin Roiland, Chri...	414849.0	23.0	NaN		Justi Roiland, Chri Parnel Spence Gramm.
4	Army of Thieves	(2021)	Action, Crime, Horror	NaN	A prequel, set before the events of Army of t...	Director: Matthias Schweighöfer   St...	NaN	NaN	NaN	Matthias Schweighöfer	Matthia Schweighöfe, Nathali Emmanue Ru.

In [5]:

```
# Dropping Gross and STARS columns:

data = data.drop(columns='Gross')
data = data.drop(columns='STARS')
data.head()
```

Out[5]:

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	VOTES	RunTime	Director	Stars
0	Blood Red Sky	(2021)	Action, Horror, Thriller	6.1	A woman with a mysterious illness is forced i...	21062.0	121.0	Peter Thorwarth	Peri Baumeister, Carl Anton Koch, Alexander ...
1	Masters of the Universe: Revelation	(2021- )	Animation, Action, Adventure	5.0	The war for Eternia begins again in what may ...	17870.0	25.0		Chris Wood, Sarah Michelle Gellar, Lena Head...
2	The Walking Dead	(2010-2022)	Drama, Horror, Thriller	8.2	Sheriff Deputy Rick Grimes wakes up from a co...	885805.0	44.0		Andrew Lincoln, Norman Reedus, Melissa McBri...
3	Rick and Morty	(2013- )	Animation, Adventure, Comedy	9.2	An animated series that follows the exploits ...	414849.0	23.0		Justin Roiland, Chris Parnell, Spencer Gramm...
4	Army of Thieves	(2021)	Action, Crime, Horror	NaN	A prequel, set before the events of Army of t...	NaN	NaN	Matthias Schweighöfer	Matthias Schweighöfer, Nathalie Emmanuel, Ru...

In [6]:

```
# Checking for null values:

data.isnull().sum()
```

Out[6]:

```
MOVIES      0
YEAR        438
GENRE        70
RATING      918
ONE-LINE     0
VOTES       918
RunTime    1350
Director      0
Stars         0
dtype: int64
```

In [7]:

```
# Filling null values:

for column in ['YEAR', 'GENRE']:
    data[column]=data[column].fillna('Not Specified')

# Filling null values with mean value of the column:

for column in ['RATING', 'VOTES', 'RunTime']:
    data[column]=data[column].fillna(int(data[column].mean()))
data
```

Out[7]:

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	VOTES	RunTime	Director	Stars
0	Blood Red Sky	(2021)	Action, Horror, Thriller	6.1	A woman with a mysterious illness is forced i...	21062.0	121.0	Peter Thorwarth	Peri Baumeister, Carl Anton Koch, Alexander ...

Chris Wood.

	MOVIES	YEAR	GENRE	RATING	ONE LINE	VOTES	RunTime	Director	Stars
1	Master Universe: Revelation	(2021- )	Action, Adventure	5.0	The war begins again in what may ...	17070.0	25.0		Michelle Gellar, Lena Head...
2	The Walking Dead	(2010-2022)	Drama, Horror, Thriller	8.2	Sheriff Deputy Rick Grimes wakes up from a co...	885805.0	44.0		Andrew Lincoln, Norman Reedus, Melissa McBri...
3	Rick and Morty	(2013- )	Animation, Adventure, Comedy	9.2	An animated series that follows the exploits ...	414849.0	23.0		Justin Roiland, Chris Parnell, Spencer Gramm...
4	Army of Thieves	(2021)	Action, Crime, Horror	6.0	A prequel, set before the events of Army of t...	21652.0	79.0	Matthias Schweighöfer	Matthias Schweighöfer, Nathalie Emmanuel, Ru...
...	...	...	...	...	...	...	...	...	...
9845	Disparu à jamais	(2021)	Crime, Drama, Mystery	6.0	Not Specified	21652.0	79.0	Juan Carlos Medina	
9901	Modern Family	(2009-2020)	Comedy, Drama, Romance	8.2	Jay must adapt to his young new wife, Gloria ...	3404.0	23.0	Jason Winer	Ed O'Neill, Sofía Vergara, Julie Bowen, Ty ...
9993	Totenfrau	(2022- )	Drama, Thriller	6.0	Not Specified	21652.0	79.0	Nicolai Rohde	Felix Klare, Romina Küper, Anna Maria Mühle, ...
9995	Arcane	(2021- )	Animation, Action, Adventure	6.0	Not Specified	21652.0	79.0		
9996	Heart of Invictus	(2022- )	Documentary, Sport	6.0	Not Specified	21652.0	79.0	Orlando von Einsiedel	

6398 rows x 9 columns

In [8]:

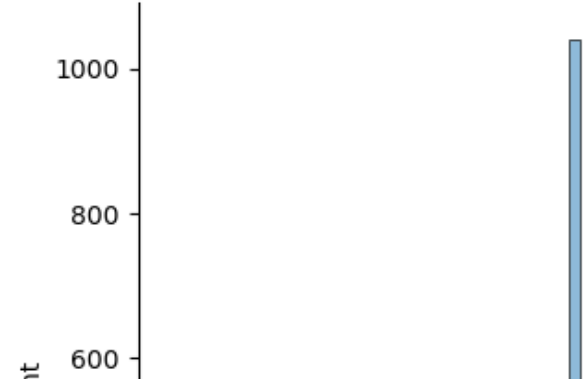
```
# Converting clean data to csv file:

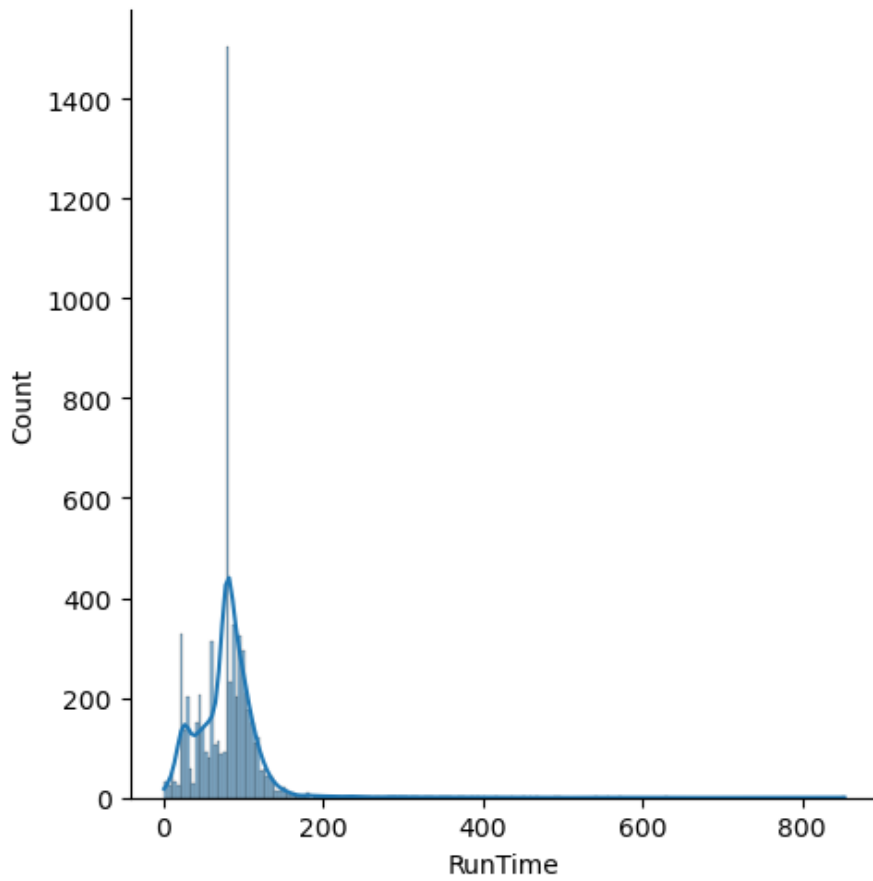
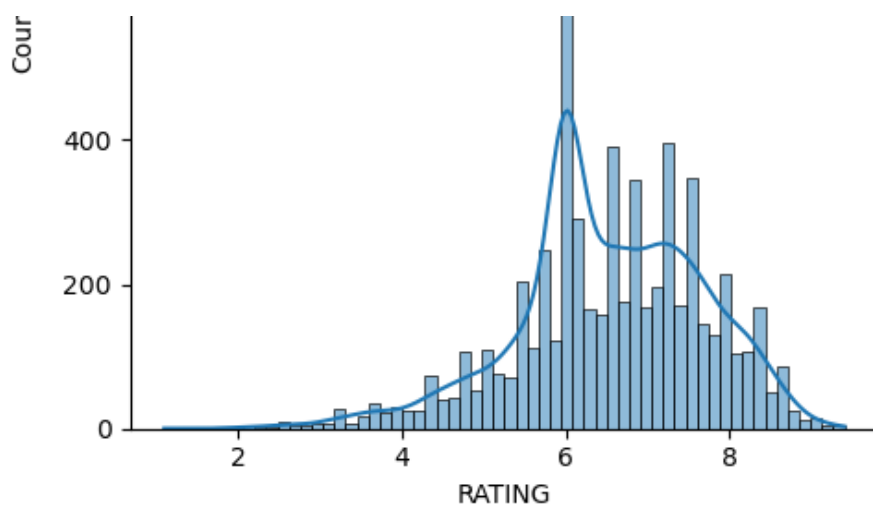
data.to_csv(r'C:\Users\samve\Documents\Data science\movies_clean.csv')
```

In [9]:

```
# DATA VISUALIZATION :-
# Distribution plots for RATING and RunTime columns:

for column in ['RATING', 'RunTime']:
    sns.displot(data[column], kind='hist', kde= True)
plt.show()
```





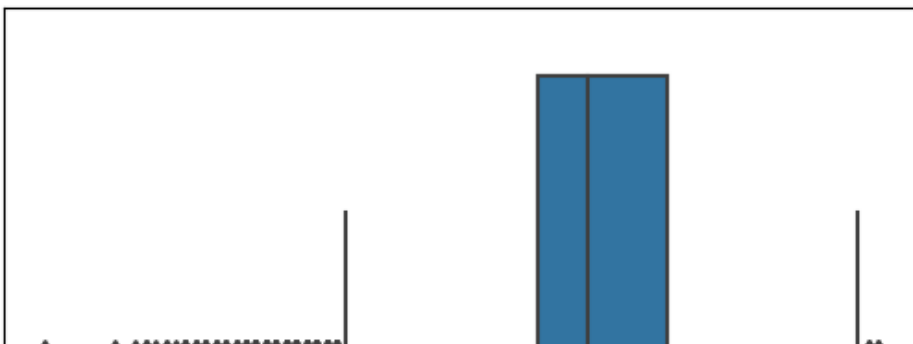
In [10]:

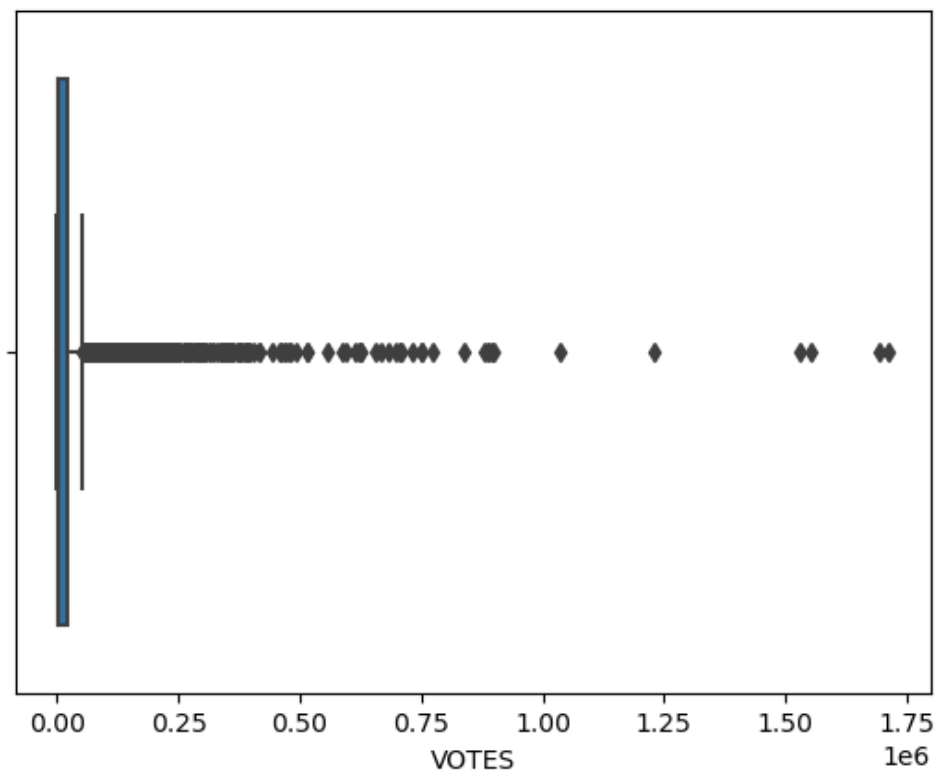
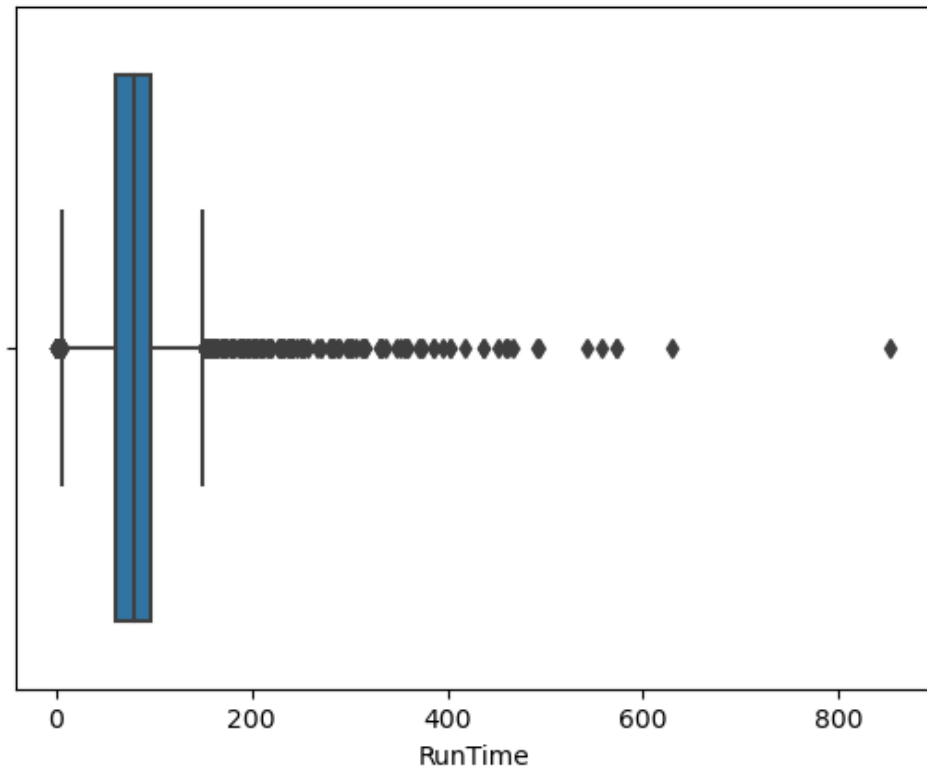
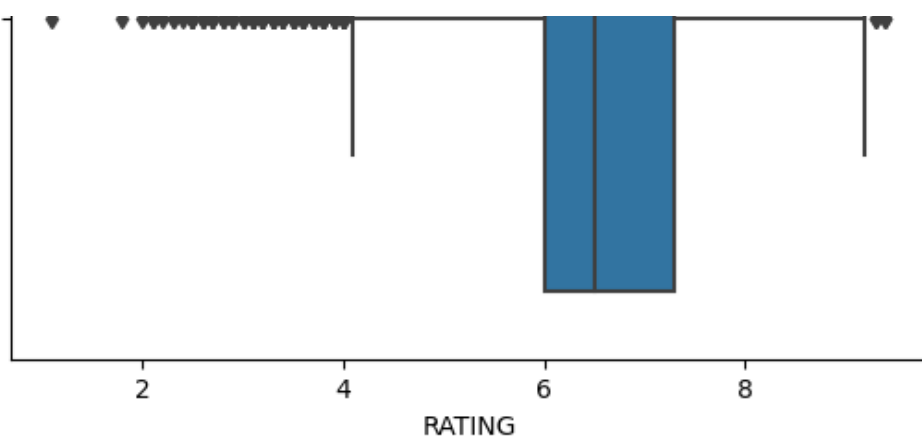
```
# Boxplots for RATING, RunTime and VOTES:
```

```
sns.boxplot(data['RATING'])
plt.show()
```

```
sns.boxplot(data['RunTime'])
plt.show()
```

```
sns.boxplot(data['VOTES'])
plt.show()
```





In [11]:

```
# Depicting Top 10:
```

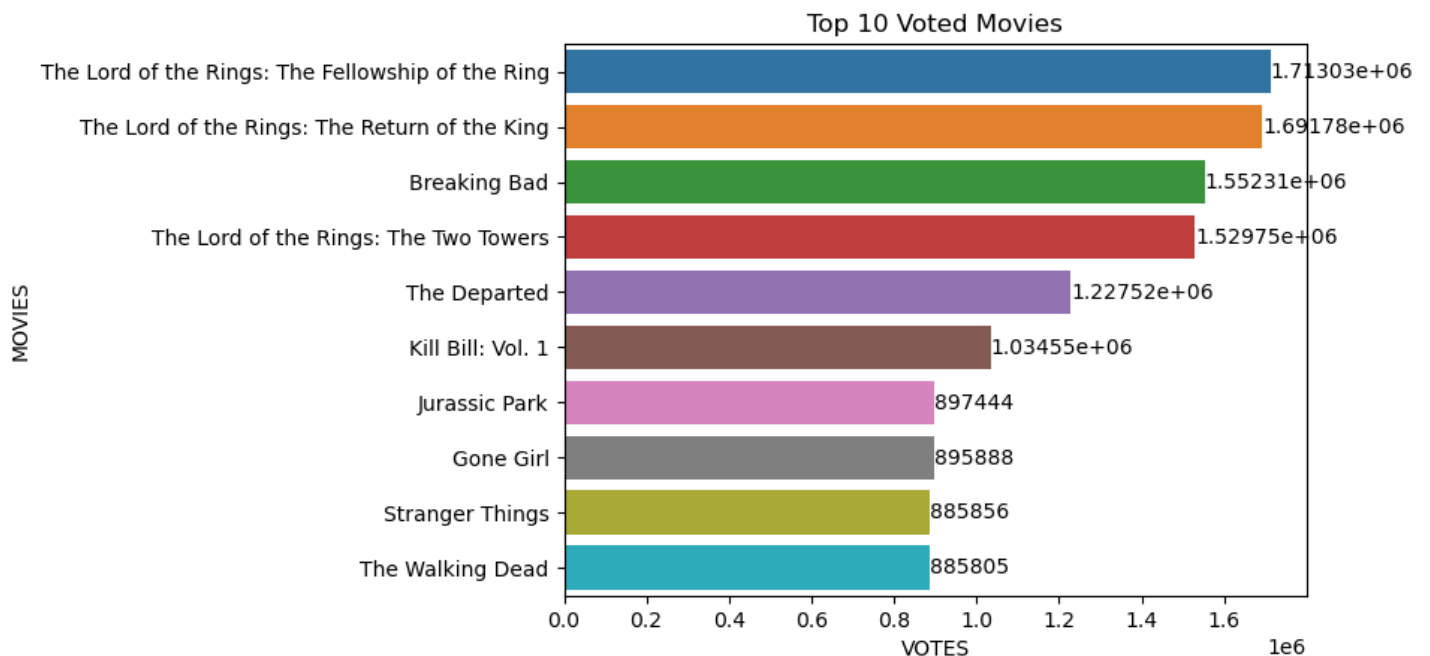
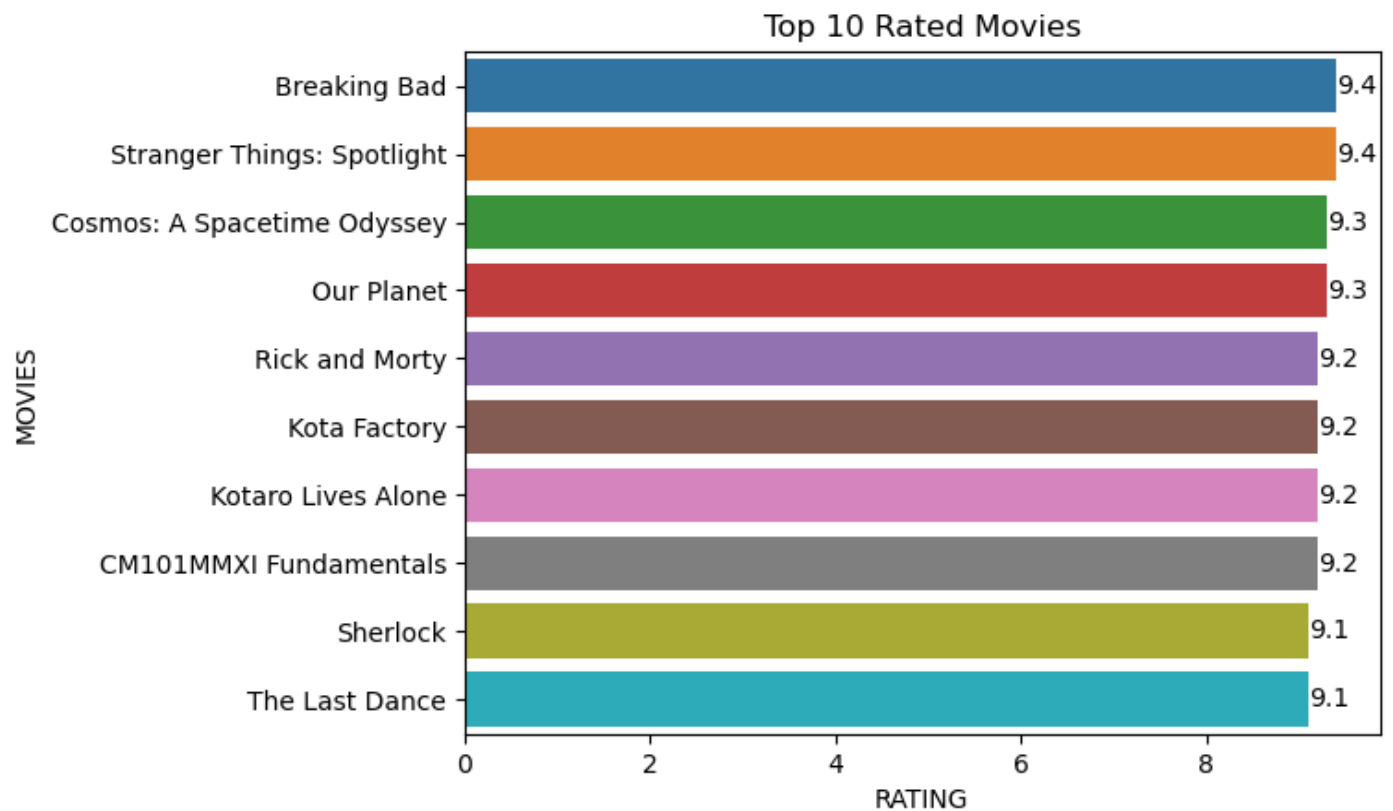
```

data_R10 = data.nlargest(10, ['RATING'])
plot= sns.barplot(data = data_R10, y= 'MOVIES', x= 'RATING')
for var in plot.containers:
    plot.bar_label(var)
plt.title('Top 10 Rated Movies')
plt.show()

data_V10 = data.nlargest(10, ['VOTES'])
plot= sns.barplot(data = data_V10, y= 'MOVIES', x= 'VOTES')
for var in plot.containers:
    plot.bar_label(var)
plt.title('Top 10 Voted Movies')
plt.show()

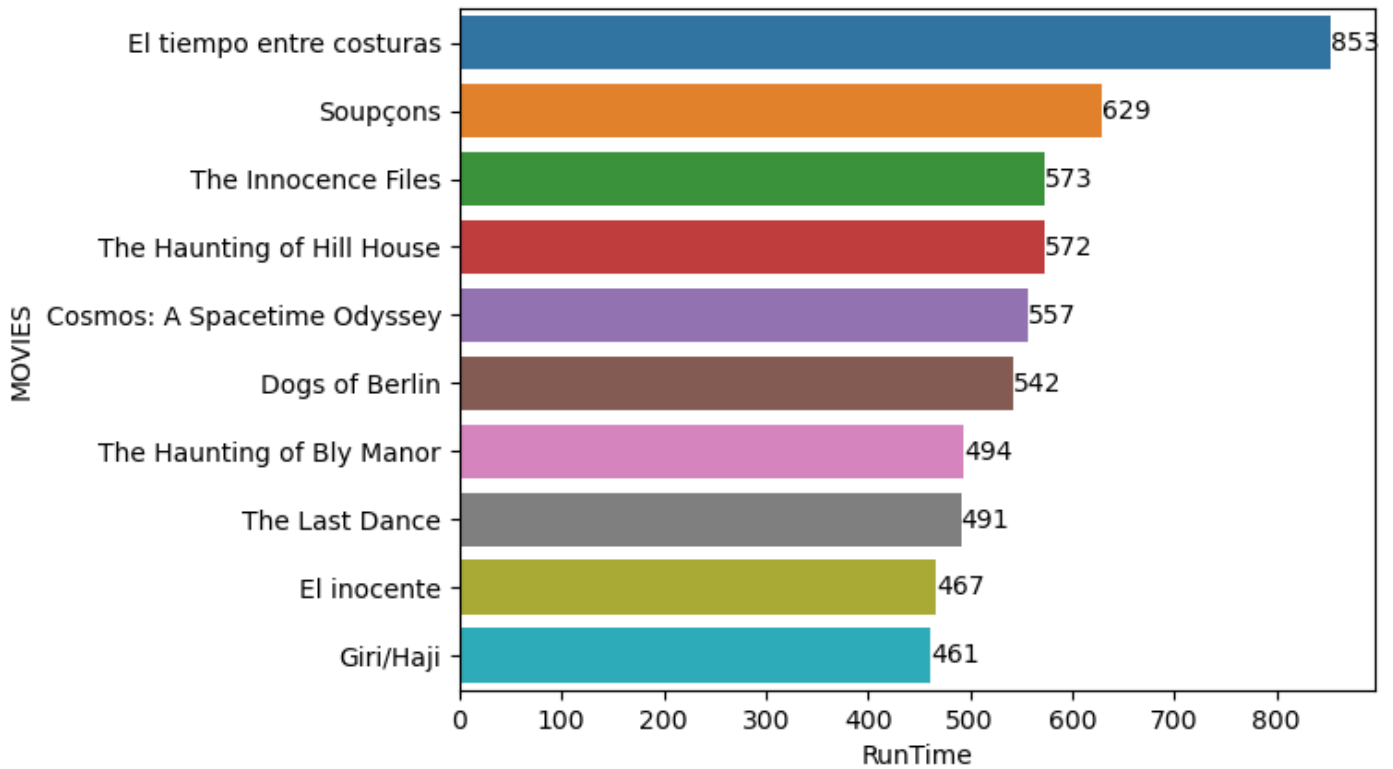
data_RT10 = data.nlargest(10, ['RunTime'])
plot= sns.barplot(data = data_RT10, y= 'MOVIES', x= 'RunTime')
for var in plot.containers:
    plot.bar_label(var)
plt.title('Top 10 Movies with Highest RunTime')
plt.show()

```





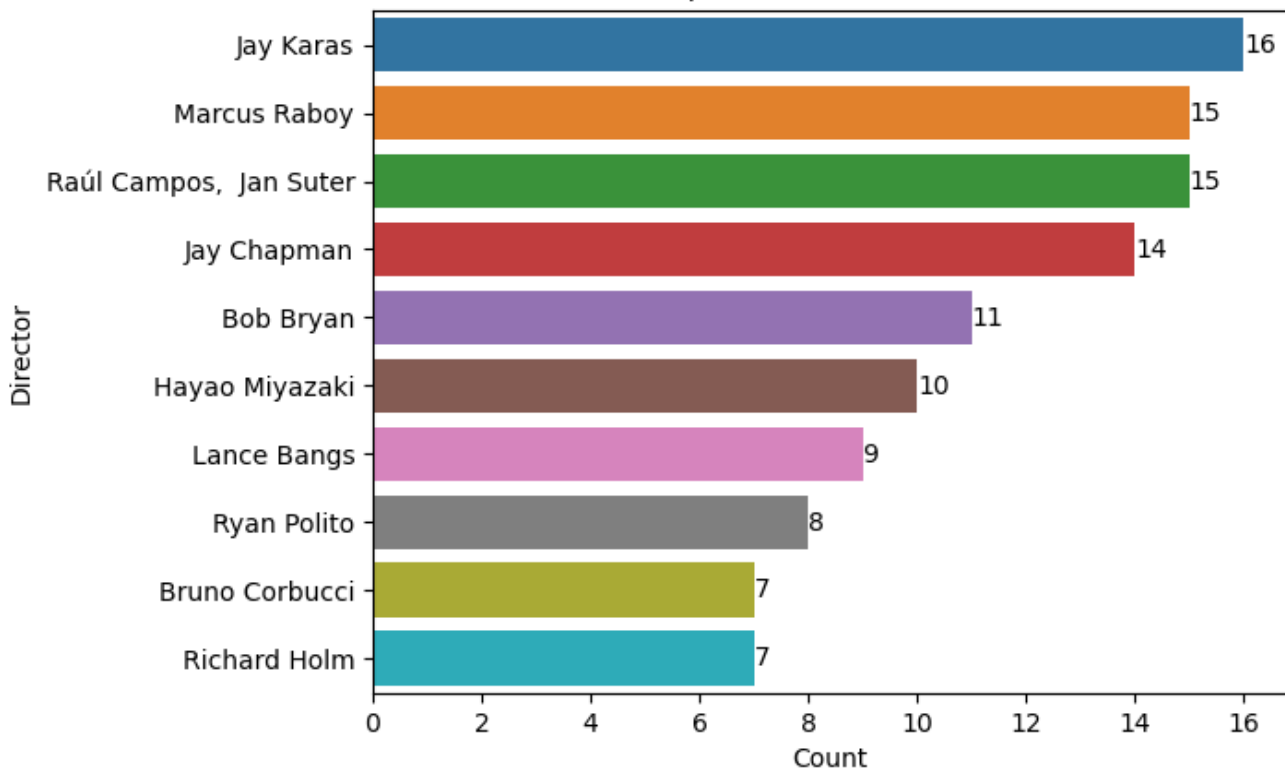
Top 10 Movies with Highest RunTime



In [12]:

```
data_direct = data[~(data['Director'] == "")]['Director'].value_counts().reset_index().head(10)
plot = sns.barplot(data = data_direct, x = 'Director', y = 'index')
plt.xlabel('Count')
plt.ylabel('Director')
plt.title('Top 10 Movie Directors')
for var in plot.containers:
    plot.bar_label(var)
plt.show()
```

Top 10 Movie Directors



In [13]:

```
from collections import Counter
```

```

genre = data['GENRE'].to_list()
genre_list = list()

for genres in genre:
    genres = genres.split(', ')
    for g in genres:
        genre_list.append(g)

data_genre = pd.DataFrame.from_dict(Counter(genre_list), orient = 'index').rename(columns = {0: 'Count'})
data_G = data_genre.sort_values(by = 'Count', ascending = False).head(10)
data_G

```

Out[13]:

	Count
<b>Animation</b>	786
<b>Comedy</b>	750
<b>Drama</b>	735
<b>Action</b>	735
<b>Drama</b>	696
<b>Comedy</b>	672
<b>Drama</b>	665
<b>Romance</b>	599
<b>Documentary</b>	553
<b>Thriller</b>	545

In [14]:

```

STARS = data[~(data['Stars'] == '')]['Stars'].to_list()
stars_list = list()

for stars in STARS:
    stars = stars.split(', ')
    for s in stars:
        stars_list.append(s)

data_stars = pd.DataFrame.from_dict(Counter(stars_list), orient = 'index').rename(columns = {0: 'Count'})
data_S = data_stars.sort_values(by = 'Count', ascending = False).head(10)
data_S

```

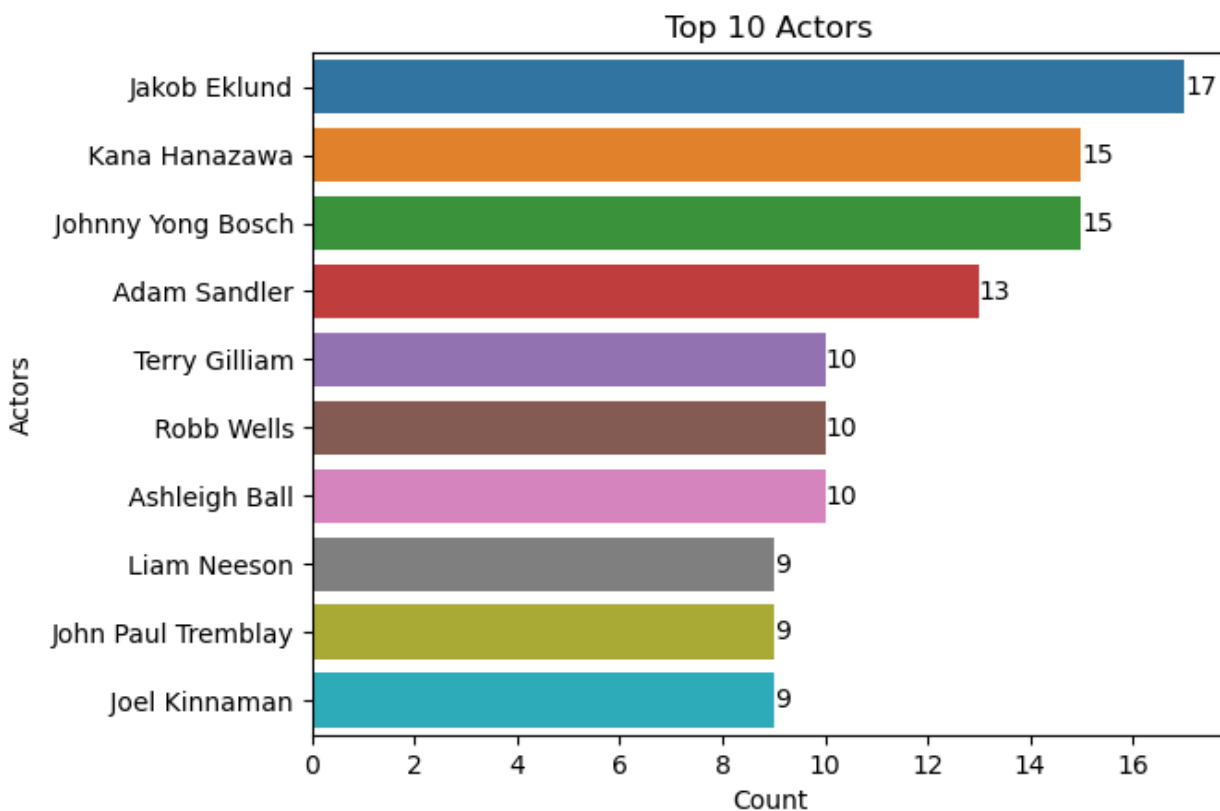
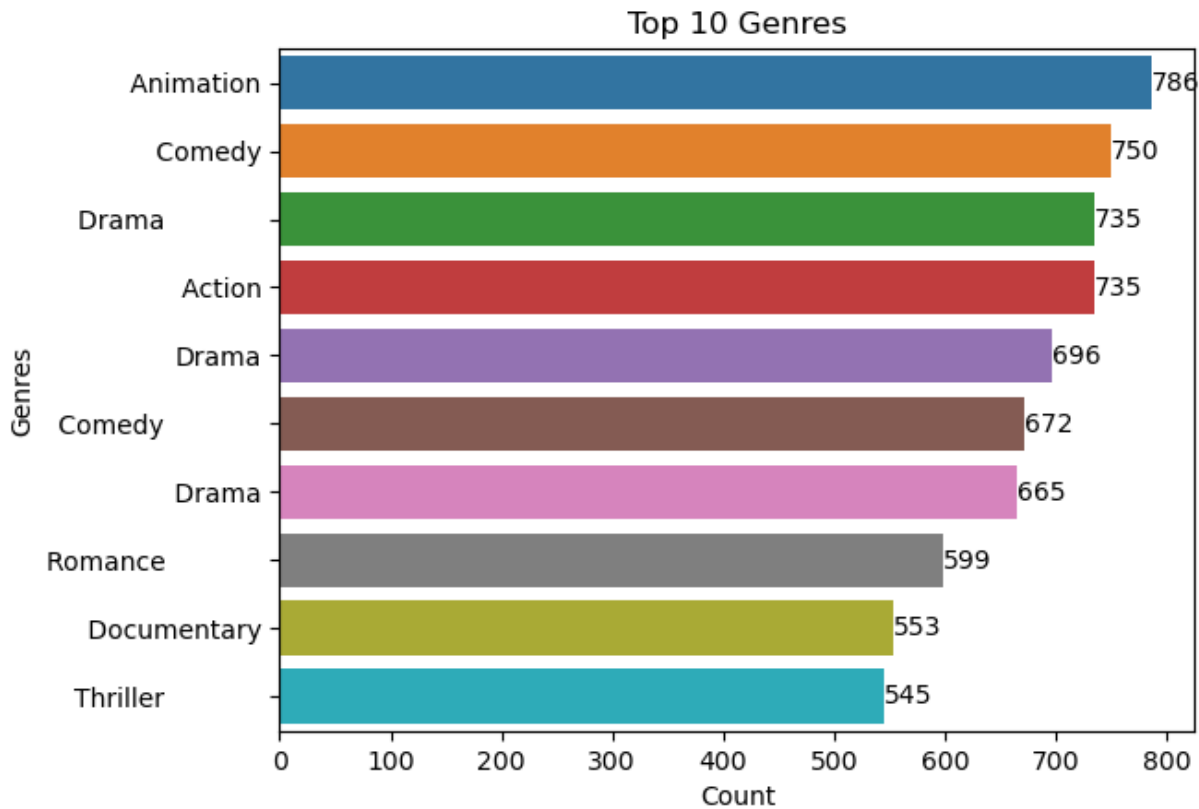
Out[14]:

	Count
<b>Jakob Eklund</b>	17
<b>Kana Hanazawa</b>	15
<b>Johnny Yong Bosch</b>	15
<b>Adam Sandler</b>	13
<b>Terry Gilliam</b>	10
<b>Robb Wells</b>	10
<b>Ashleigh Ball</b>	10
<b>Liam Neeson</b>	9
<b>John Paul Tremblay</b>	9
<b>Joel Kinnaman</b>	9

In [15]:

```
plot = sns.barplot(data = data_G, x = 'Count', y = data_G.index)
plt.ylabel('Genres')
for var in plot.containers:
    plot.bar_label(var)
plt.title('Top 10 Genres')
plt.show()

plot = sns.barplot(data = data_S, x = 'Count', y = data_S.index)
plt.ylabel('Actors')
for var in plot.containers:
    plot.bar_label(var)
plt.title('Top 10 Actors')
plt.show()
```



In [16]:

```
# Features

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity

features = ['GENRE', 'ONE-LINE', 'Director', 'Stars']

data['combined_features'] = data['GENRE'] + " " + data['ONE-LINE'] + " " + data['Director'] + " " + data['Stars']
cv = CountVectorizer()
count_matrix = cv.fit_transform(data['combined_features'])
cosine_sim = cosine_similarity(count_matrix)
```

In [17]:

```
# Movie Recommendation Function

def movie_recommendation(mov, sim_num = 5):

    user_choice = mov

    try:
        ref_index = data[data['MOVIES'].str.contains(user_choice, case = False)].index[0]

        similar_movies = list(enumerate(cosine_sim[ref_index]))

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

        print('\nRecomended Movies for [{}]' .format(user_choice))
        print('-'*(24 + len(user_choice)))

        for i, element in enumerate(sorted_similar_movies):
            similar_movie_id = element[0]
            similar_movie_title = data['MOVIES'].iloc[similar_movie_id]
            s_score = element[1]
            print('{:40} -> {:.3f}' .format(similar_movie_title, s_score))

            if i > sim_num:
                break
    except IndexError:
        print("\n[{}] is not in our database!" .format(user_choice))
        print("We couldn't recommend anything...Sorry...")
```

In [20]:

```
# Searching for a movie using a keyword

def movie_with(key):

    keyword = key

    print("Movie with keyword: [{}]" .format(keyword))

    for i, mov in enumerate(data[data['MOVIES'].str.contains(keyword)][['MOVIES']].to_list()):
        print("{} {} " .format(i+1, mov))
```

In [22]:

```
# Running keyword function

movie_with('land')
```

Movie with keyword: [land]

- 1) Outlander
- 2) Heartland
- 3) Homeland
- 4) Kong: Skull Island

5) This Is England  
6) Fantasy Island  
7) Into the Badlands  
8) Yakusoku no Neverland  
9) Alice in Borderland  
10) Love Island  
11) Tidelands  
12) Young Wallander  
13) Portlandia  
14) Deadman Wonderland  
15) Grenseland  
16) Love Island Australia  
17) Blue Mountain State: The Rise of Thadland  
18) Slumberland  
19) Skylanders Academy  
20) Zombieland Saga  
21) Sunderland 'Til I Die  
22) Bon Bini Holland  
23) Mysterious Island  
24) Larva Island  
25) Strong Island  
26) Zoolander: Super Model  
27) Ever After High: Way Too Wonderland  
28) Dreamland  
29) Nerdland  
30) Alice in Wonderland  
31) Le voyage au Groenland  
32) The Larva Island Movie  
33) Pardon My Icelandic  
34) Skull Island  
35) Judah Friedlander: America is the Greatest Country in the United States  
36) Holland: Natuur in de Delta  
37) The Island  
38) Moshe Kasher: Live in Oakland  
39) Highland: Thailand's Marijuana Awakening  
40) Skylanders Academy: Happy Birthday to You!  
41) Brasslands  
42) Strongland  
43) Southland's Most Wanted  
44) Mr. D: Southland to Japan  
45) One Piece: Enter Chopper at the Winter Island

In [24]:

```
# Running recommendation function
```

```
movie_recommendation('The Walking Dead')
```

Recommended Movies for [The Walking Dead]

```
----->
Gerald's Game          -> 0.367
Hija única              -> 0.358
Vampire in the Garden   -> 0.349
Exception               -> 0.345
The Queen               -> 0.345
Prison Break            -> 0.343
Sonora                  -> 0.342
```

In [ ]: