

In [1]:

```
import pandas as pd
import numpy as np
```

Extracting features of 2018 movies from Wikipedia

In [2]:

```
link = "https://en.wikipedia.org/wiki/List_of_American_films_of_2018"
df1 = pd.read_html(link, header=0)[2]
df2 = pd.read_html(link, header=0)[3]
df3 = pd.read_html(link, header=0)[4]
df4 = pd.read_html(link, header=0)[5]
```

In [3]:

```
df = df1.append(df2.append(df3.append(df4, ignore_index=True), ignore_index=True), ignore_index=True)
```

In [4]:

df

Out[4]:

	Opening	Opening.1	Title	Production company	Cast and crew	Ref.
0	JANUARY	5	Insidious: The Last Key	Universal Pictures / Blumhouse Productions / S...	Adam Robitel (director); Leigh Whannell (scree...	[2]
1	JANUARY	5	The Strange Ones	Vertical Entertainment	Lauren Wolkstein (director); Christopher Radcl...	[3]
2	JANUARY	5	Stratton	Momentum Pictures	Simon West (director); Duncan Falconer, Warren...	[4]
3	JANUARY	10	Sweet Country	Samuel Goldwyn Films	Warwick Thornton (director); David Tranter, St...	[5]
4	JANUARY	12	The Commuter	Lionsgate / StudioCanal / The Picture Company	Jaume Collet-Serra (director); Byron Willinger...	[6]
...
263	DECEMBER	25	Holmes & Watson	Columbia Pictures / Gary Sanchez Productions	Etan Cohen (director/screenplay); Will Ferrell...	[162]
264	DECEMBER	25	Vice	Annapurna Pictures / Plan B Entertainment	Adam McKay (director/screenplay); Christian Ba...	[136]
265	DECEMBER	25	On the Basis of Sex	Focus Features	Mimi Leder (director); Daniel Stiepleman (scre...	[223]
266	DECEMBER	25	Destroyer	Annapurna Pictures	Karyn Kusama (director); Phil Hay, Matt Manfre...	[256]
267	DECEMBER	28	Black Mirror: Bandersnatch	Netflix	David Slade (director); Charlie Brooker (scree...	[257]

268 rows x 6 columns

In [5]:

```
from tmdbv3api import TMDb
import json
import requests
tmdb = TMDb()
tmdb.api_key = ''
```

In [6]:

```
In [6]:
```

```
from tmdbv3api import Movie
tmdb_movie = Movie()
def get_genre(x):
    genres = []
    result = tmdb_movie.search(x)
    movie_id = result[0].id
    response = requests.get('https://api.themoviedb.org/3/movie/{}?api_key={}'.format(movie_id, tmdb.api_key))
    data_json = response.json()
    if data_json['genres']:
        genre_str = " "
        for i in range(0, len(data_json['genres'])):
            genres.append(data_json['genres'][i]['name'])
        return genre_str.join(genres)
    else:
        np.NaN
```

```
In [7]:
```

```
df['genres'] = df['Title'].map(lambda x: get_genre(str(x)))
```

```
In [8]:
```

```
df
```

```
Out[8]:
```

	Opening	Opening.1	Title	Production company	Cast and crew	Ref.	genres
0	JANUARY	5	Insidious: The Last Key	Universal Pictures / Blumhouse Productions / S...	Adam Robitel (director); Leigh Whannell (scree...	[2]	Mystery Horror Thriller
1	JANUARY	5	The Strange Ones	Vertical Entertainment	Lauren Wolkstein (director); Christopher Radcl...	[3]	Thriller Drama
2	JANUARY	5	Stratton	Momentum Pictures	Simon West (director); Duncan Falconer, Warren...	[4]	Action Thriller
3	JANUARY	10	Sweet Country	Samuel Goldwyn Films	Warwick Thornton (director); David Tranter, St...	[5]	Drama History Western
4	JANUARY	12	The Commuter	Lionsgate / StudioCanal / The Picture Company	Jaume Collet-Serra (director); Byron Willinger...	[6]	Action Thriller
...
263	DECEMBER	25	Holmes & Watson	Columbia Pictures / Gary Sanchez Productions	Etan Cohen (director/screenplay); Will Ferrell...	[162]	Mystery Adventure Comedy Crime
264	DECEMBER	25	Vice	Annapurna Pictures / Plan B Entertainment	Adam McKay (director/screenplay); Christian Ba...	[136]	Thriller Science Fiction Action Adventure
265	DECEMBER	25	On the Basis of Sex	Focus Features	Mimi Leder (director); Daniel Stiepleman (scre...	[223]	Drama History
266	DECEMBER	25	Destroyer	Annapurna Pictures	Karyn Kusama (director); Phil Hay, Matt Manfre...	[256]	Thriller Crime Drama Action
267	DECEMBER	28	Black Mirror: Bandersnatch	Netflix	David Slade (director); Charlie Brooker (scre...	[257]	Science Fiction Mystery Drama Thriller TV Movie

268 rows x 7 columns

```
In [9]:
```

```
df_2018 = df[['Title', 'Cast and crew', 'genres']]
```

In [10]:

```
df_2018
```

Out[10]:

	Title	Cast and crew	genres
0	Insidious: The Last Key	Adam Robitel (director); Leigh Whannell (scree...	Mystery Horror Thriller
1	The Strange Ones	Lauren Wolkstein (director); Christopher Radcl...	Thriller Drama
2	Stratton	Simon West (director); Duncan Falconer, Warren...	Action Thriller
3	Sweet Country	Warwick Thornton (director); David Tranter, St...	Drama History Western
4	The Commuter	Jaume Collet-Serra (director); Byron Willinger...	Action Thriller
...
263	Holmes & Watson	Etan Cohen (director/screenplay); Will Ferrell...	Mystery Adventure Comedy Crime
264	Vice	Adam McKay (director/screenplay); Christian Ba...	Thriller Science Fiction Action Adventure
265	On the Basis of Sex	Mimi Leder (director); Daniel Stiepleman (scre...	Drama History
266	Destroyer	Karyn Kusama (director); Phil Hay, Matt Manfre...	Thriller Crime Drama Action
267	Black Mirror: Bandersnatch	David Slade (director); Charlie Brooker (scree...	Science Fiction Mystery Drama Thriller TV Movie

268 rows x 3 columns

In [11]:

```
def get_director(x):  
    if " (director)" in x:  
        return x.split(" (director)") [0]  
    elif " (directors)" in x:  
        return x.split(" (directors)") [0]  
    else:  
        return x.split(" (director/screenplay)") [0]
```

In [12]:

```
df_2018['director_name'] = df_2018['Cast and crew'].map(lambda x: get_director(x))
```

<ipython-input-12-4838efec492a>: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

```
df_2018['director_name'] = df_2018['Cast and crew'].map(lambda x: get_director(x))
```

In [13]:

```
def get_actor1(x):  
    return ((x.split("screenplay); ")[-1]).split(", ")[0])
```

In [14]:

```
df_2018['actor_1_name'] = df_2018['Cast and crew'].map(lambda x: get_actor1(x))
```

<ipython-input-14-94b1a1569c5d>: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

```
df_2018['actor_1_name'] = df_2018['Cast and crew'].map(lambda x: get_actor1(x))
```

In [15]:

```
def get_actor2(x):
    if len((x.split("screenplay"); ")[-1]).split(", ")) < 2:
        return np.NaN
    else:
        return ((x.split("screenplay"); ")[-1]).split(", ")[1])
```

In [16]:

```
df_2018['actor_2_name'] = df_2018['Cast and crew'].map(lambda x: get_actor2(x))
```

<ipython-input-16-75f2be6b779a>: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

```
df_2018['actor_2_name'] = df_2018['Cast and crew'].map(lambda x: get_actor2(x))
```

In [17]:

```
def get_actor3(x):
    if len((x.split("screenplay"); ")[-1]).split(", ")) < 3:
        return np.NaN
    else:
        return ((x.split("screenplay"); ")[-1]).split(", ")[2])
```

In [18]:

```
df_2018['actor_3_name'] = df_2018['Cast and crew'].map(lambda x: get_actor3(x))
```

In [19]:

```
df_2018
```

Out[19]:

	Title	Cast and crew	genres	director_name	actor_1_name	actor_2_name	actor_3_name
0	Insidious: The Last Key	Adam Robitel (director); Leigh Whannell (scre...	Mystery Horror Thriller	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell
1	The Strange Ones	Lauren Wolkstein (director); Christopher Radcl...	Thriller Drama	Lauren Wolkstein	Alex Pettyfer	James Freedson-Jackson	Emily Althaus
2	Stratton	Simon West (director); Duncan Falconer, Warren...	Action Thriller	Simon West	Dominic Cooper	Austin Stowell	Gemma Chan
3	Sweet Country	Warwick Thornton (director); David Tranter, St...	Drama History Western	Warwick Thornton	Bryan Brown	Sam Neill	NaN
4	The Commuter	Jaume Collet-Serra (director); Byron Willinger...	Action Thriller	Jaume Collet-Serra	Liam Neeson	Vera Farmiga	Patrick Wilson
...
263	Holmes & Watson	Etan Cohen (director/screenplay); Will Ferrell...	Mystery Adventure Comedy Crime	Etan Cohen	Will Ferrell	John C. Reilly	Rebecca Hall
264	Vice	Adam McKay (director/screenplay); Christian Ba...	Thriller Science Fiction Action Adventure	Adam McKay	Christian Bale	Amy Adams	Steve Carell
265	On the Basis of Sex	Mimi Leder (director); Daniel Stiepleman (scre...	Drama History	Mimi Leder	Felicity Jones	Armie Hammer	Justin Theroux
266	Destroyer	Karyn Kusama (director); Phil Hay, Matt Monf...	Thriller Crime Drama Action	Karyn Kusama	Nicole Kidman	Sebastian Stan	Toby Kebbell

	Title	Cast and crew	genres	director_name	actor_1_name	actor_2_name	actor_3_name
267	Black Mirror: Bandersnatch	David Slade (director); Charlie Brooker (scree...	Science Fiction Mystery Drama Thriller TV Movie	David Slade	Fionn Whitehead	Will Poulter	Asim Chaudhry

268 rows x 7 columns

In [20]:

```
df_2018 = df_2018.rename(columns={'Title': 'movie_title'})
```

In [21]:

```
new_df18 = df_2018.loc[:, ['director_name', 'actor_1_name', 'actor_2_name', 'actor_3_name', 'genres', 'movie_title']]
```

In [22]:

```
new_df18
```

Out[22]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title
0	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell	Mystery Horror Thriller	Insidious: The Last Key
1	Lauren Wolkstein	Alex Pettyfer	James Freedson-Jackson	Emily Althaus	Thriller Drama	The Strange Ones
2	Simon West	Dominic Cooper	Austin Stowell	Gemma Chan	Action Thriller	Stratton
3	Warwick Thornton	Bryan Brown	Sam Neill	NaN	Drama History Western	Sweet Country
4	Jaume Collet-Serra	Liam Neeson	Vera Farmiga	Patrick Wilson	Action Thriller	The Commuter
...
263	Etan Cohen	Will Ferrell	John C. Reilly	Rebecca Hall	Mystery Adventure Comedy Crime	Holmes & Watson
264	Adam McKay	Christian Bale	Amy Adams	Steve Carell	Thriller Science Fiction Action Adventure	Vice
265	Mimi Leder	Felicity Jones	Armie Hammer	Justin Theroux	Drama History	On the Basis of Sex
266	Karyn Kusama	Nicole Kidman	Sebastian Stan	Toby Kebbell	Thriller Crime Drama Action	Destroyer
267	David Slade	Fionn Whitehead	Will Poulter	Asim Chaudhry	Science Fiction Mystery Drama Thriller TV Movie	Black Mirror: Bandersnatch

268 rows x 6 columns

In [23]:

```
new_df18['actor_2_name'] = new_df18['actor_2_name'].replace(np.nan, 'unknown')
new_df18['actor_3_name'] = new_df18['actor_3_name'].replace(np.nan, 'unknown')
```

In [24]:

```
new_df18['movie_title'] = new_df18['movie_title'].str.lower()
```

In [25]:

```
new_df18['comb'] = new_df18['actor_1_name'] + ' ' + new_df18['actor_2_name'] + ' ' + new_df18['actor_3_name'] + ' ' + new_df18['director_name'] + ' ' + new_df18['genres']
```

In [26]:

```
new_df18
```

Out [26]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
0	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell	Mystery Horror Thriller	insidious: the last key	Lin Shaye Angus Sampson Leigh Whannell Adam Ro...
1	Lauren Wolkstein	Alex Pettyfer	James Freedson-Jackson	Emily Althaus	Thriller Drama	the strange ones	Alex Pettyfer James Freedson-Jackson Emily Alt...
2	Simon West	Dominic Cooper	Austin Stowell	Gemma Chan	Action Thriller	stratton	Dominic Cooper Austin Stowell Gemma Chan Simon...
3	Warwick Thornton	Bryan Brown	Sam Neill	unknown	Drama History Western	sweet country	Bryan Brown Sam Neill unknown Warwick Thornton...
4	Jaume Collet-Serra	Liam Neeson	Vera Farmiga	Patrick Wilson	Action Thriller	the commuter	Liam Neeson Vera Farmiga Patrick Wilson Jaume ...
...
263	Etan Cohen	Will Ferrell	John C. Reilly	Rebecca Hall	Mystery Adventure Comedy Crime	holmes & watson	Will Ferrell John C. Reilly Rebecca Hall Etan ...
264	Adam McKay	Christian Bale	Amy Adams	Steve Carell	Thriller Science Fiction Action Adventure	vice	Christian Bale Amy Adams Steve Carell Adam McK...
265	Mimi Leder	Felicity Jones	Armie Hammer	Justin Theroux	Drama History	on the basis of sex	Felicity Jones Armie Hammer Justin Theroux Mim...
266	Karyn Kusama	Nicole Kidman	Sebastian Stan	Toby Kebbell	Thriller Crime Drama Action	destroyer	Nicole Kidman Sebastian Stan Toby Kebbell Kary...
267	David Slade	Fionn Whitehead	Will Poulter	Asim Chaudhry	Science Fiction Mystery Drama Thriller TV Movie	black mirror: bandersnatch	Fionn Whitehead Will Poulter Asim Chaudhry Dav...

268 rows x 7 columns

Extracting features of 2019 movies from Wikipedia

In [27]:

```
link = "https://en.wikipedia.org/wiki/List_of_American_films_of_2019"
df1 = pd.read_html(link, header=0)[2]
df2 = pd.read_html(link, header=0)[3]
df3 = pd.read_html(link, header=0)[4]
df4 = pd.read_html(link, header=0)[5]
```

In [28]:

```
df = df1.append(df2.append(df3.append(df4, ignore_index=True), ignore_index=True), ignore_index=True)
```

In [29]:

```
df
```

Out [29]:

	Opening	Opening.1	Title	Production company	Cast and crew	Ref.
0	JANUARY	4	Escape Room	Columbia Pictures	Adam Robitel (director); Bragi F. Schut, Maria...	[2]

1	Opening	4	Rust Creek	Production Company	Jen McGowan (director); Julie Lipson (screenplay); Donal...	Ref
2	JANUARY	4	American Hangman	Hangman Justice Productions	Wilson Coneybeare (director/screenplay); Donal...	[4]
3	JANUARY	11	A Dog's Way Home	Columbia Pictures	Charles Martin Smith (director); W. Bruce Came...	[5]
4	JANUARY	11	The Upside	STX Entertainment	Neil Burger (director); Jon Hartmere (screenpl...	[6]
...
236	DECEMBER	25	Little Women	Columbia Pictures / Regency Enterprises	Greta Gerwig (director/screenplay); Saoirse Ro...	[222]
237	DECEMBER	25	1917	Universal Pictures / DreamWorks Pictures / Ent...	Sam Mendes (director/screenplay); Krysty Wilso...	[223]
238	DECEMBER	25	Just Mercy	Warner Bros. Pictures / Participant Media	Destin Daniel Cretton (director/screenplay), A...	[224]
239	DECEMBER	27	Clemency	Neon	Chinonye Chukwu (director/screenplay); Alfre W...	[225]
240	DECEMBER	27	Apparition	Vertical Pictures	Waymon Boone (director/screenplay); Mena Suvar...	[226]

241 rows x 6 columns

In [30]:

```
df['genres'] = df['Title'].map(lambda x: get_genre(str(x)))
```

In [31]:

```
df_2019 = df[['Title', 'Cast and crew', 'genres']]
```

In [32]:

```
df_2019
```

Out[32]:

	Title	Cast and crew	genres
0	Escape Room	Adam Robitel (director); Bragi F. Schut, Maria...	Thriller Action Mystery Horror
1	Rust Creek	Jen McGowan (director); Julie Lipson (screenpl...	Thriller Drama
2	American Hangman	Wilson Coneybeare (director/screenplay); Donal...	Thriller
3	A Dog's Way Home	Charles Martin Smith (director); W. Bruce Came...	Drama Adventure Family
4	The Upside	Neil Burger (director); Jon Hartmere (screenpl...	Comedy Drama
...
236	Little Women	Greta Gerwig (director/screenplay); Saoirse Ro...	Drama Romance
237	1917	Sam Mendes (director/screenplay); Krysty Wilso...	War Drama Action History
238	Just Mercy	Destin Daniel Cretton (director/screenplay), A...	Drama Crime
239	Clemency	Chinonye Chukwu (director/screenplay); Alfre W...	Drama
240	Apparition	Waymon Boone (director/screenplay); Mena Suvar...	Horror Thriller

241 rows x 3 columns

In [33]:

```
def get_director(x):
    if " (director)" in x:
        return x.split(" (director)") [0]
    elif " (directors)" in x:
        return x.split(" (directors)") [0]
```

```
else:
    return x.split(" (director/screenplay)") [0]
```

In [34]:

```
df_2019['director_name'] = df_2019['Cast and crew'].map(lambda x: get_director(str(x)))
```

<ipython-input-34-a433a4faa834>: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

```
df_2019['director_name'] = df_2019['Cast and crew'].map(lambda x: get_director(str(x)))
```

In [35]:

```
def get_actor1(x):
    return ((x.split("screenplay; ") [-1]).split(", ") [0])
```

In [36]:

```
df_2019['actor_1_name'] = df_2019['Cast and crew'].map(lambda x: get_actor1(x))
```

<ipython-input-36-eee71c5586b8>: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

```
df_2019['actor_1_name'] = df_2019['Cast and crew'].map(lambda x: get_actor1(x))
```

In [37]:

```
def get_actor2(x):
    if len((x.split("screenplay; ") [-1]).split(", ")) < 2:
        return np.NaN
    else:
        return ((x.split("screenplay; ") [-1]).split(", ") [1])
```

In [38]:

```
df_2019['actor_2_name'] = df_2019['Cast and crew'].map(lambda x: get_actor2(x))
```

<ipython-input-38-48d0aad78c0>: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

```
df_2019['actor_2_name'] = df_2019['Cast and crew'].map(lambda x: get_actor2(x))
```

In [39]:

```
def get_actor3(x):
    if len((x.split("screenplay; ") [-1]).split(", ")) < 3:
        return np.NaN
    else:
        return ((x.split("screenplay; ") [-1]).split(", ") [2])
```

In [40]:

```
df_2019['actor_3_name'] = df_2019['Cast and crew'].map(lambda x: get_actor3(x))
```

In [41]:

```
df_2019 = df_2019.rename(columns={'Title': 'movie_title'})
```

In [42]:

```
new df19 = df 2019.loc[:, ['director name', 'actor 1 name', 'actor 2 name', 'actor 3 name', 'q
```



```
enres', 'movie_title']]
```

In [43]:

```
new_df19['actor_2_name'] = new_df19['actor_2_name'].replace(np.nan, 'unknown')
new_df19['actor_3_name'] = new_df19['actor_3_name'].replace(np.nan, 'unknown')
```

In [44]:

```
new_df19['movie_title'] = new_df19['movie_title'].str.lower()
```

In [45]:

```
new_df19['comb'] = new_df19['actor_1_name'] + ' ' + new_df19['actor_2_name'] + ' ' + new_
df19['actor_3_name'] + ' ' + new_df19['director_name'] + ' ' + new_df19['genres']
```

In [46]:

```
new_df19
```

Out[46]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
0	Adam Robitel	Taylor Russell	Logan Miller	Deborah Ann Woll	Thriller Action Mystery Horror	escape room	Taylor Russell Logan Miller Deborah Ann Woll A...
1	Jen McGowan	Hermione Corfield	Jay Paulson	Sean O'Bryan	Thriller Drama	rust creek	Hermione Corfield Jay Paulson Sean O'Bryan Jen...
2	Wilson Coneybeare	Donald Sutherland	Vincent Kartheiser	Oliver Dennis	Thriller	american hangman	Donald Sutherland Vincent Kartheiser Oliver De...
3	Charles Martin Smith	Bryce Dallas Howard	Edward James Olmos	Alexandra Shipp	Drama Adventure Family	a dog's way home	Bryce Dallas Howard Edward James Olmos Alexand...
4	Neil Burger	Bryan Cranston	Kevin Hart	Nicole Kidman	Comedy Drama	the upside	Bryan Cranston Kevin Hart Nicole Kidman Neil B...
...
236	Greta Gerwig	Saoirse Ronan	Emma Watson	Florence Pugh	Drama Romance	little women	Saoirse Ronan Emma Watson Florence Pugh Greta ...
237	Sam Mendes	George MacKay	Dean-Charles Chapman	Mark Strong	War Drama Action History	1917	George MacKay Dean-Charles Chapman Mark Strong...
238	Destin Daniel Cretton	Michael B. Jordan	Jamie Foxx	Brie Larson	Drama Crime	just mercy	Michael B. Jordan Jamie Foxx Brie Larson Desti...
239	Chinonye Chukwu	Alfre Woodard	Wendell Pierce	Aldis Hodge	Drama	clemency	Alfre Woodard Wendell Pierce Aldis Hodge Chino...
240	Waymon Boone	Mena Suvari	Kevin Pollak	unknown	Horror Thriller	apparition	Mena Suvari Kevin Pollak unknown Waymon Boone ...

241 rows x 7 columns

In [47]:

```
my_df = new_df18.append(new_df19, ignore_index=True)
```

In [50]:

```
my_df
```

Out[50]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
0	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell	Mystery Horror Thriller	insidious: the last key	Lin Shaye Angus Sampson Leigh Whannell Adam Ro...

1	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	Alex Pettyfer	James Freedson-Jackson	Emily Alt...
2	Simon West	Dominic Cooper	Austin Stowell	Gemma Chan	Action Thriller	stratton	Dominic Cooper	Austin Stowell	Gemma Chan Simon...
3	Warwick Thornton	Bryan Brown	Sam Neill	unknown	Drama History Western	sweet country	Bryan Brown	Sam Neill	unknown Warwick Thornton...
4	Jaume Collet-Serra	Liam Neeson	Vera Farmiga	Patrick Wilson	Action Thriller	the commuter	Liam Neeson	Vera Farmiga	Patrick Wilson Jaume ...
...
504	Greta Gerwig	Saoirse Ronan	Emma Watson	Florence Pugh	Drama Romance	little women	Saoirse Ronan	Emma Watson	Florence Pugh Greta ...
505	Sam Mendes	George MacKay	Dean-Charles Chapman	Mark Strong	War Drama Action History	1917	George MacKay	Dean-Charles Chapman	Mark Strong...
506	Destin Daniel Cretton	Michael B. Jordan	Jamie Foxx	Brie Larson	Drama Crime	just mercy	Michael B. Jordan	Jamie Foxx	Brie Larson Desti...
507	Chinonye Chukwu	Alfre Woodard	Wendell Pierce	Aldis Hodge	Drama	clemency	Alfre Woodard	Wendell Pierce	Aldis Hodge Chino...
508	Waymon Boone	Mena Suvari	Kevin Pollak	unknown	Horror Thriller	apparition	Mena Suvari	Kevin Pollak	unknown Waymon Boone ...

509 rows × 7 columns

In [51]:

```
old_df = pd.read_csv('../datasets/new_data.csv')
```

In [52]:

```
old_df
```

Out[52]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb		
0	James Cameron	CCH Pounder	Joel David Moore	Wes Studi	Action Adventure Fantasy Sci-Fi	avatar	CCH Pounder	Joel David Moore	Wes Studi James C...
1	Gore Verbinski	Johnny Depp	Orlando Bloom	Jack Davenport	Action Adventure Fantasy	pirates of the caribbean: at world's end	Johnny Depp	Orlando Bloom	Jack Davenport Gore ...
2	Sam Mendes	Christoph Waltz	Rory Kinnear	Stephanie Sigman	Action Adventure Thriller	spectre	Christoph Waltz	Rory Kinnear	Stephanie Sigman ...
3	Christopher Nolan	Tom Hardy	Christian Bale	Joseph Gordon-Levitt	Action Thriller	the dark knight rises	Tom Hardy	Christian Bale	Joseph Gordon-Levitt ...
4	Doug Walker	Doug Walker	Rob Walker	unknown	Documentary	star wars: episode vii - the force awakens ...	Doug Walker	Rob Walker	unknown Doug Walker Doc...
...
5359	Jim Strouse	Jessica Williams	Chris O'Dowd	Keith Stanfield	Romance Comedy	the incredible jessica james	Jessica Williams	Chris O'Dowd	Keith Stanfield ...
5360	Farhad Mann	Adelaide Kane	Benjamin Hollingsworth	Jean Louisa Kelly	Romance	can't buy my love	Adelaide Kane	Benjamin Hollingsworth	Jean Loui...
5361	Hannaleena Hauru	Inka Haapamäki	Rosa Honkonen	Tiitus Rantala	Romance Comedy	thick lashes of lauri mäntyvaara	Inka Haapamäki	Rosa Honkonen	Tiitus Rantala Ha...

5362	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
	Rosenbaum	Lou Diamond Phillips	Wallace Shawn	Gina Holden	Crime Comedy	new recruit	Lou Diamond Phillips Wallace Shawn Gina Holden...
5363	Ravi Udyawar	Sridevi Kapoor	Sajal Ali	Akshaye Khanna	Crime Drama Thriller	mom	Sridevi Kapoor Sajal Ali Akshaye Khanna Ravi U...

5364 rows x 7 columns

In [53]:

```
final_df = old_df.append(my_df, ignore_index=True)
```

In [54]:

```
final_df
```

Out[54]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
0	James Cameron	CCH Pounder	Joel David Moore	Wes Studi	Action Adventure Fantasy Sci-Fi	avatar	CCH Pounder Joel David Moore Wes Studi James C...
1	Gore Verbinski	Johnny Depp	Orlando Bloom	Jack Davenport	Action Adventure Fantasy	pirates of the caribbean: at world's end	Johnny Depp Orlando Bloom Jack Davenport Gore ...
2	Sam Mendes	Christoph Waltz	Rory Kinnear	Stephanie Sigman	Action Adventure Thriller	spectre	Christoph Waltz Rory Kinnear Stephanie Sigman ...
3	Christopher Nolan	Tom Hardy	Christian Bale	Joseph Gordon-Levitt	Action Thriller	the dark knight rises	Tom Hardy Christian Bale Joseph Gordon-Levitt ...
4	Doug Walker	Doug Walker	Rob Walker	unknown	Documentary	star wars: episode vii - the force awakens ...	Doug Walker Rob Walker unknown Doug Walker Doc...
...
5868	Greta Gerwig	Saoirse Ronan	Emma Watson	Florence Pugh	Drama Romance	little women	Saoirse Ronan Emma Watson Florence Pugh Greta ...
5869	Sam Mendes	George MacKay	Dean-Charles Chapman	Mark Strong	War Drama Action History	1917	George MacKay Dean-Charles Chapman Mark Strong...
5870	Destin Daniel Cretton	Michael B. Jordan	Jamie Foxx	Brie Larson	Drama Crime	just mercy	Michael B. Jordan Jamie Foxx Brie Larson Desti...
5871	Chinonye Chukwu	Alfre Woodard	Wendell Pierce	Aldis Hodge	Drama	clemency	Alfre Woodard Wendell Pierce Aldis Hodge Chino...
5872	Waymon Boone	Mena Suvari	Kevin Pollak	unknown	Horror Thriller	apparition	Mena Suvari Kevin Pollak unknown Waymon Boone ...

5873 rows x 7 columns

In [55]:

```
final_df.isna().sum()
```

Out[55]:

```
director_name    0
actor_1_name     0
actor_2_name     0
actor_3_name     0
genres           1
```

```
genres          1  
movie_title     0  
comb           4  
dtype: int64
```

In [56]:

```
final_df = final_df.dropna(how='any')
```

In [57]:

```
final_df.isna().sum()
```

Out[57]:

```
director_name    0  
actor_1_name     0  
actor_2_name     0  
actor_3_name     0  
genres           0  
movie_title      0  
comb             0  
dtype: int64
```

In [58]:

```
final_df.to_csv('../datasets/final_data.csv', index=False)
```