

```
In [82]: import pandas as pd
import numpy as np
```

## Movie 2018

```
In [83]: 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 [84]: df = df1.append(df2.append(df3.append(df4, ignore_index=True), ignore_index=True), i
```

In [85]: df

Out[85]:

	Opening	Opening.1	Title	Production company	Cast and crew	
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
...	...	...	...	...	...	..
265	DECEMBER	25	Holmes & Watson	Columbia Pictures / Gary Sanchez Productions	Etan Cohen (director/screenplay); Will Ferrell...	Na
266	DECEMBER	25	Vice	Annapurna Pictures / Plan B Entertainment	Adam McKay (director/screenplay); Christian Ba...	Na
267	DECEMBER	25	On the Basis of Sex	Focus Features	Mimi Leder (director); Daniel Stiepleman (scre...	Na
268	DECEMBER	25	Destroyer	Annapurna Pictures	Karyn Kusama (director); Phil Hay, Matt Manfre...	Na
269	DECEMBER	28	Black Mirror: Bandersnatch	Netflix	David Slade (director); Charlie Brooker (scree...	Na

270 rows x 7 columns



```
In [6]: !pip install tmdbv3api
```

Collecting tmdbv3api

Downloading tmdbv3api-1.7.6-py2.py3-none-any.whl (17 kB)

Requirement already satisfied: requests in c:\users\voan\anaconda3\lib\site-packages (from tmdbv3api) (2.25.1)

Requirement already satisfied: chardet<5,>=3.0.2 in c:\users\voan\anaconda3\lib\site-packages (from requests->tmdbv3api) (4.0.0)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\voan\anaconda3\lib\site-packages (from requests->tmdbv3api) (2020.12.5)

Requirement already satisfied: idna<3,>=2.5 in c:\users\voan\anaconda3\lib\site-packages (from requests->tmdbv3api) (2.10)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\voan\anaconda3\lib\site-packages (from requests->tmdbv3api) (1.26.4)

Installing collected packages: tmdbv3api

Successfully installed tmdbv3api-1.7.6

```
In [86]: from tmdbv3api import TMDb
import json
import requests
tmdb = TMDb()
tmdb.api_key = 'b7e7aeaa608257af93a032205bfada32'
```

```
In [87]: 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:
        return np.NaN
```

```
In [88]: df['genres'] = df['Title'].map(lambda x: get_genre(str(x)))
```

In [89]:

df

Out[89]:

	Opening	Opening.1	Title	Production company	Cast and crew	.mw-parser-output .tooltip-dotted{border-bottom:1px dotted;cursor:help}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
...	...	...	...	...	...	..
265	DECEMBER	25	Holmes & Watson	Columbia Pictures / Gary Sanchez Productions	Etan Cohen (director/screenplay); Will Ferrell...	Na
266	DECEMBER	25	Vice	Annapurna Pictures / Plan B Entertainment	Adam McKay (director/screenplay); Christian Ba...	Na
267	DECEMBER	25	On the Basis of Sex	Focus Features	Mimi Leder (director); Daniel Stiepleman (scre...	Na
268	DECEMBER	25	Destroyer	Annapurna Pictures	Karyn Kusama (director); Phil Hay, Matt Manfre...	Na
269	DECEMBER	28	Black Mirror: Bandersnatch	Netflix	David Slade (director); Charlie Brooker (scree...	Na

270 rows × 8 columns

```
In [90]: df_2018 = df[['Title', 'Cast and crew', 'genres']]
```

```
In [91]: df_2018
```

```
Out[91]:
```

	Title	Cast and crew	genres
0	Insidious: The Last Key	Adam Robitel (director); Leigh Whannell (scree...	Horror Mystery 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
...	...	...	...
265	Holmes & Watson	Etan Cohen (director/screenplay); Will Ferrell...	Mystery Adventure Comedy Crime
266	Vice	Adam McKay (director/screenplay); Christian Ba...	Thriller Science Fiction Action Adventure
267	On the Basis of Sex	Mimi Leder (director); Daniel Stiepleman (scre...	Drama History
268	Destroyer	Karyn Kusama (director); Phil Hay, Matt Manfre...	Thriller Crime Drama Action
269	Black Mirror: Bandersnatch	David Slade (director); Charlie Brooker (scree...	Science Fiction Mystery Drama Thriller TV Movie

270 rows × 3 columns

```
In [92]: #Lấy ra tên đạo diễn
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 [93]: df_2018['director_name'] = df_2018['Cast and crew'].map(lambda x: get_director(x))
```

<ipython-input-93-4838efec492a>: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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](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 [94]: #Lấy ra 3 tên diễn viên
def get_actor1(x):
    return ((x.split("screenplay); ")[-1]).split(", ")[0])
```

```
In [95]: df_2018['actor_1_name'] = df_2018['Cast and crew'].map(lambda x: get_actor1(x))
```

<ipython-input-95-94b1a1569c5d>: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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](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 [96]: def get_actor2(x):
    if len((x.split("screenplay); ")[-1]).split(", ")) < 2:
        return np.NaN
    else:
        return ((x.split("screenplay); ")[-1]).split(", ")[1])
```

```
In [97]: df_2018['actor_2_name'] = df_2018['Cast and crew'].map(lambda x: get_actor2(x))
```

<ipython-input-97-75f2be6b779a>: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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](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 [98]: def get_actor3(x):  
         if len((x.split("screenplay; ")[-1]).split(", ")) < 3:  
             return np.NaN  
         else:  
             return ((x.split("screenplay; ")[-1]).split(", ")[2])
```

```
In [99]: df_2018['actor_3_name'] = df_2018['Cast and crew'].map(lambda x: get_actor3(x))
```

```
<ipython-input-99-c6969dc4d5e1>: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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

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

In [100]: df\_2018

Out[100]:

	Title	Cast and crew	genres	director_name	actor_1_name	actor_2_name	act
0	Insidious: The Last Key	Adam Robitel (director); Leigh Whannell (scree...	Horror Mystery Thriller	Adam Robitel	Lin Shaye	Angus Sampson	
1	The Strange Ones	Lauren Wolkstein (director); Christopher Radcl...	Thriller Drama	Lauren Wolkstein	Alex Pettyfer	James Freedson-Jackson	Er
2	Stratton	Simon West (director); Duncan Falconer, Warren...	Action Thriller	Simon West	Dominic Cooper	Austin Stowell	Ge
3	Sweet Country	Warwick Thornton (director); David Tranter, St...	Drama History Western	Warwick Thornton	Bryan Brown	Sam Neill	
4	The Commuter	Jaume Collet-Serra (director); Byron Willinger...	Action Thriller	Jaume Collet-Serra	Liam Neeson	Vera Farmiga	Pa
...	...	...	...	...	...	...	...
265	Holmes & Watson	Etan Cohen (director/screenplay); Will Ferrell...	Mystery Adventure Comedy Crime	Etan Cohen	Will Ferrell	John C. Reilly	R
266	Vice	Adam McKay (director/screenplay); Christian Ba...	Thriller Science Fiction Action Adventure	Adam McKay	Christian Bale	Amy Adams	S
267	On the Basis of Sex	Mimi Leder (director); Daniel Stiepleman (scree...	Drama History	Mimi Leder	Felicity Jones	Armie Hammer	Jus
268	Destroyer	Karyn Kusama (director); Phil Hay, Matt Manfre...	Thriller Crime Drama Action	Karyn Kusama	Nicole Kidman	Sebastian Stan	T
269	Black Mirror: Bandersnatch	David Slade (director); Charlie Brooker (scree...	Science Fiction Mystery Drama Thriller TV Movie	David Slade	Fionn Whitehead	Will Poulter	

270 rows × 7 columns



In [101]: `#Đổi cột title = movie_title`  
`df_2018 = df_2018.rename(columns={'Title': 'movie_title'})`



```
In [102]: new_df18 = df_2018.loc[:, ['director_name', 'actor_1_name', 'actor_2_name', 'actor_3_
```

```
In [103]: new_df18
```

```
Out[103]:
```

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title
0	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell	Horror Mystery 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
...	...	...	...	...	...	...
265	Etan Cohen	Will Ferrell	John C. Reilly	Rebecca Hall	Mystery Adventure Comedy Crime	Holmes & Watson
266	Adam McKay	Christian Bale	Amy Adams	Steve Carell	Thriller Science Fiction Action Adventure	Vice
267	Mimi Leder	Felicity Jones	Armie Hammer	Justin Theroux	Drama History	On the Basis of Sex
268	Karyn Kusama	Nicole Kidman	Sebastian Stan	Toby Kebbell	Thriller Crime Drama Action	Destroyer
269	David Slade	Fionn Whitehead	Will Poulter	Asim Chaudhry	Science Fiction Mystery Drama Thriller TV Movie	Black Mirror: Bandersnatch

270 rows × 6 columns

```
In [104]: #Tiền xử lý dữ liệu thay thế giá trị na thành unknow
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 [105]: new_df18['movie_title'] = new_df18['movie_title'].str.lower()
```

```
In [106]: new_df18['comb'] = new_df18['actor_1_name'] + ' ' + new_df18['actor_2_name'] + ' '
```

In [107]:

```
new_df18
```

Out[107]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comt
0	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell	Horror Mystery Thriller	insidious: the last key	Lin Shaye Angus Sampson Leigh Whannell Adan 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 Coope Austin Stowel 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 ..
...	...	...	...	...	...	...	..
265	Etan Cohen	Will Ferrell	John C. Reilly	Rebecca Hall	Mystery Adventure Comedy Crime	holmes & watson	Will Ferrell John C Reilly Rebecca Hall Etan ..
266	Adam McKay	Christian Bale	Amy Adams	Steve Carell	Thriller Science Fiction Action Adventure	vice	Christian Bale Amy Adams Steve Carell Adan McK..
267	Mimi Leder	Felicity Jones	Armie Hammer	Justin Theroux	Drama History	on the basis of sex	Felicity Jones Armie Hammer Justin Theroux Mim..

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	coml
268	Karyn Kusama	Nicole Kidman	Sebastian Stan	Toby Kebbell	Thriller Crime Drama Action	destroyer	Nicole Kidman Sebastian Stan Toby Kebbell Karyn..
269	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..

270 rows × 7 columns



## Movie 2019

```
In [108]: link = "https://en.wikipedia.org/wiki/List_of_American_films_of_2019"
df1 = pd.read_html(link, header=0)[3]
df2 = pd.read_html(link, header=0)[4]
df3 = pd.read_html(link, header=0)[5]
df4 = pd.read_html(link, header=0)[6]
```

```
In [109]: df = df1.append(df2.append(df3.append(df4, ignore_index=True), ignore_index=True), i
```

In [110]: df

Out[110]:

	Opening	Opening.1	Title	Production company	Cast and crew	Ref.	Unnamed: 0	Wikime Comm has me relate Film: the Uni Stai 20
0	APRIL	5.0	Shazam!	Warner Bros. Pictures / New Line Cinema / DC F...	David F. Sandberg (director); Henry Gayden (sc...	[51]	NaN	ℳ
1	APRIL	5.0	Pet Sematary	Paramount Pictures	Kevin Kolsch, Dennis Widmyer (directors); Jeff...	[52]	NaN	ℳ
2	APRIL	5.0	The Best of Enemies	STX Entertainment	Robin Bissell (director/screenplay); Taraji P...	[53]	NaN	ℳ
3	APRIL	5.0	High Life	A24 / Arte / ZDF / Canal+ / Ciné+	Claire Denis (director/screenplay); Jean-Pol F...	[54]	NaN	ℳ
4	APRIL	5.0	Unicorn Store	Netflix	Brie Larson (director); Samantha McIntyre (scr...	NaN	NaN	ℳ
...	...	...	...	...	...	...	...	...
188	DECEMBER	25.0	Spies in Disguise	20th Century Fox / Blue Sky Studios / Chernin ...	Nick Bruno, Troy Quane (directors); Brad Copel...	[132]	NaN	ℳ
189	DECEMBER	25.0	Little Women	Columbia Pictures / Regency Enterprises	Greta Gerwig (director/screenplay); Saoirse Ro...	[221]	NaN	ℳ
190	DECEMBER	25.0	1917	Universal Pictures / DreamWorks Pictures / Ent...	Sam Mendes (director/screenplay); Krysty Wilso...	[222]	NaN	ℳ
191	DECEMBER	25.0	Just Mercy	Warner Bros. Pictures / Participant Media	Destin Daniel Cretton (director/screenplay), A...	[223]	NaN	ℳ
192	DECEMBER	27.0	Clemency	Neon	Chinonye Chukwu (director/screenplay); Alfre W...	[224]	NaN	ℳ

193 rows × 8 columns

```
In [111]: df['genres'] = df['Title'].map(lambda x: get_genre(str(x)))
```

```
In [112]: df_2019 = df[['Title', 'Cast and crew', 'genres']]
```

```
In [113]: df_2019
```

Out[113]:

	Title	Cast and crew	genres
0	Shazam!	David F. Sandberg (director); Henry Gayden (sc...	Action Comedy Fantasy
1	Pet Sematary	Kevin Kolsch, Dennis Widmyer (directors); Jeff...	Horror
2	The Best of Enemies	Robin Bissell (director/screenplay); Taraji P....	Drama History
3	High Life	Claire Denis (director/screenplay); Jean-Pol F...	Science Fiction Drama Mystery Thriller
4	Unicorn Store	Brie Larson (director); Samantha McIntyre (scr...	Fantasy Drama Comedy
...	...	...	...
188	Spies in Disguise	Nick Bruno, Troy Quane (directors); Brad Copel...	Animation Action Adventure Comedy Family
189	Little Women	Greta Gerwig (director/screenplay); Saoirse Ro...	Drama Romance
190	1917	Sam Mendes (director/screenplay); Krysty Wilso...	War Drama Action Thriller
191	Just Mercy	Destin Daniel Cretton (director/screenplay), A...	Drama Crime History
192	Clemency	Chinonye Chukwu (director/screenplay); Alfre W...	Drama

193 rows × 3 columns

```
In [114]: df_2019['director_name'] = df_2019['Cast and crew'].map(lambda x: get_director(st
```

```
<ipython-input-114-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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](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 [115]: df_2019['actor_1_name'] = df_2019['Cast and crew'].map(lambda x: get_actor1(x))
```

```
<ipython-input-115-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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](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 [116]: df_2019['actor_2_name'] = df_2019['Cast and crew'].map(lambda x: get_actor2(x))
```

```
<ipython-input-116-48d0aad78c0>: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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](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 [117]: df_2019['actor_3_name'] = df_2019['Cast and crew'].map(lambda x: get_actor3(x))
```

```
<ipython-input-117-54baf96f0c44>: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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

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

```
In [118]: df_2019 = df_2019.rename(columns={'Title': 'movie_title'})
```

```
In [119]: new_df19 = df_2019.loc[:, ['director_name', 'actor_1_name', 'actor_2_name', 'actor_3_
```

```
In [120]:
```

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

```
In [122]: new_df19['comb'] = new_df19['actor_1_name'] + ' ' + new_df19['actor_2_name'] + ' '
```

In [129]: new\_df19

Out[129]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
0	David F. Sandberg	Zachary Levi	Asher Angel	Mark Strong	Action Comedy Fantasy	shazam!	Zachary Levi Asher Angel Mark Strong David F. ...
1	Kevin Kolsch, Dennis Widmyer	Jason Clarke	Amy Seimetz	John Lithgow	Horror	pet sematary	Jason Clarke Amy Seimetz John Lithgow Kevin Ko...
2	Robin Bissell	Taraji P. Henson	Sam Rockwell	unknown	Drama History	the best of enemies	Taraji P. Henson Sam Rockwell unknown Robin Bi...
3	Claire Denis	Robert Pattinson	Juliette Binoche	André Benjamin	Science Fiction Drama Mystery Thriller	high life	Robert Pattinson Juliette Binoche André Benjam...
4	Brie Larson	Brie Larson	Samuel L. Jackson	Joan Cusack	Fantasy Drama Comedy	unicorn store	Brie Larson Samuel L. Jackson Joan Cusack Brie...
...	...	...	...	...	...	...	...
188	Nick Bruno, Troy Quane	Will Smith	Tom Holland	Rashida Jones	Animation Action Adventure Comedy Family	spies in disguise	Will Smith Tom Holland Rashida Jones Nick Brun...
189	Greta Gerwig	Saoirse Ronan	Emma Watson	Florence Pugh	Drama Romance	little women	Saoirse Ronan Emma Watson Florence Pugh Greta ...

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
190	Sam Mendes	George Mackay	Dean-Charles Chapman	Mark Strong	War Drama Action Thriller	1917	George Mackay Dean- Charles Chapman Mark Strong...
191	Destin Daniel Cretton	Michael B. Jordan	Jamie Foxx	Brie Larson	Drama Crime History	just mercy	Michael B. Jordan Jamie Foxx Brie Larson Desti...
192	Chinonye Chukwu	Alfre Woodard	Wendell Pierce	Aldis Hodge	Drama	clemency	Alfre Woodard Wendell Pierce Aldis Hodge Chino...

193 rows × 7 columns

```
In [130]: my_df = new_df18.append(new_df19, ignore_index=True)
```



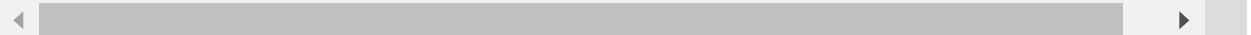
In [131]: my\_df

Out[131]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	cast
0	Adam Robitel	Lin Shaye	Angus Sampson	Leigh Whannell	Horror Mystery Thriller	insidious: the last key	Lin Shaye Angus Sampson Leigh Whannell Adrian Pasdar Richard Brake
1	Lauren Wolkstein	Alex Pettyfer	James Freedson-Jackson	Emily Althaus	Thriller Drama	the strange ones	Alex Pettyfer James Freedson-Jackson Emily Althaus
2	Simon West	Dominic Cooper	Austin Stowell	Gemma Chan	Action Thriller	stratton	Dominic Cooper Austin Stowell Gemma Chan Simon West
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 Collet-Serra
...	...	...	...	...	...	...	...
458	Nick Bruno, Troy Quane	Will Smith	Tom Holland	Rashida Jones	Animation Action Adventure Comedy Family	spies in disguise	Will Smith Tom Holland Rashida Jones Nick Bruno Troy Quane
459	Greta Gerwig	Saoirse Ronan	Emma Watson	Florence Pugh	Drama Romance	little women	Saoirse Ronan Emma Watson Florence Pugh Greta Gerwig
460	Sam Mendes	George Mackay	Dean-Charles Chapman	Mark Strong	War Drama Action Thriller	1917	George Mackay Dean-Charles Chapman Mark Strong

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	cor
461	Destin Daniel Cretton	Michael B. Jordan	Jamie Foxx	Brie Larson	Drama Crime History	just mercy	Michael Jorc Jar Foxx B Lars Des
462	Chinonye Chukwu	Alfre Woodard	Wendell Pierce	Aldis Hodge	Drama	clemency	Al Wood: Wenc Pier Al Hoc Chin

463 rows × 7 columns



```
In [134]: old_df = pd.read_csv('movie.csv')
```

In [135]: old\_df

Out[135]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
0	John Lasseter	Tom Hanks	Tim Allen	Don Rickles	Animation Comedy Family	toy story	Tom Hanks Tim Allen Don Rickles John Lasseter ...
1	Joe Johnston	Robin Williams	Jonathan Hyde	Kirsten Dunst	Adventure Fantasy Family	jumanji	Robin Williams Jonathan Hyde Kirsten Dunst Joe...
2	Howard Deutch	Walter Matthau	Jack Lemmon	Ann-Margret	Romance Comedy	grumpier old men	Walter Matthau Jack Lemmon Ann-Margret Howard ...
3	Forest Whitaker	Whitney Houston	Angela Bassett	Loretta Devine	Comedy Drama Romance	waiting to exhale	Whitney Houston Angela Bassett Loretta Devine ...
4	Charles Shyer	Steve Martin	Diane Keaton	Martin Short	Comedy	father of the bride part ii	Steve Martin Diane Keaton Martin Short Charles...
...	...	...	...	...	...	...	...
39196	Ben Rock	Monty Bane	Lucy Butler	David Grammer	Horror	the burkittsville 7	Monty Bane Lucy Butler David Grammer Ben Rock ...
39197	Aaron Osborne	Lisa Boyle	Kena Land	Zaneta Polard	Sci-Fi	caged heat 3000	Lisa Boyle Kena Land Zaneta Polard Aaron Osbor...

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	comb
39198	John Irvin	Patrick Bergin	Uma Thurman	David Morrissey	Drama Action Romance	robin hood	Patrick Bergin Uma Thurman David Morrissey Joh...
39199	Lav Diaz	Angel Aquino	Perry Dizon	Hazel Orencio	Drama	century of birthing	Angel Aquino Perry Dizon Hazel Orencio Lav Dia...
39200	Mark L. Lester	Erika Eleniak	Adam Baldwin	Julie du Page	Action Drama Thriller	betrayal	Erika Eleniak Adam Baldwin Julie du Page Mark ...

39201 rows × 7 columns



```
In [136]: final_df = old_df.append(my_df,ignore_index=True)
```

In [137]: final\_df

Out[137]:

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	cc
0	John Lasseter	Tom Hanks	Tim Allen	Don Rickles	Animation Comedy Family	toy story	Ha Tim A Ric J Lass
1	Joe Johnston	Robin Williams	Jonathan Hyde	Kirsten Dunst	Adventure Fantasy Family	jumanji	R Willi Jonai F Kir D J
2	Howard Deutch	Walter Matthau	Jack Lemmon	Ann-Margret	Romance Comedy	grumpier old men	W Mat , Lemi / Mar Hov
3	Forest Whitaker	Whitney Houston	Angela Bassett	Loretta Devine	Comedy Drama Romance	waiting to exhale	Whi Hou An Bas Loi Devir
4	Charles Shyer	Steve Martin	Diane Keaton	Martin Short	Comedy	father of the bride part ii	S M D Ke M S Charl
...	...	...	...	...	...	...	
39659	Nick Bruno, Troy Quane	Will Smith	Tom Holland	Rashida Jones	Animation Action Adventure Comedy Family	spies in disguise	S Hol Ras Jc Bri
39660	Greta Gerwig	Saoirse Ronan	Emma Watson	Florence Pugh	Drama Romance	little women	Sac Rc Er Wa Flore F Gret

	director_name	actor_1_name	actor_2_name	actor_3_name	genres	movie_title	cc
39661	Sam Mendes	George MacKay	Dean-Charles Chapman	Mark Strong	War Drama Action Thriller	1917	Ge Mac De Ch Chap M Stro
39662	Destin Daniel Cretton	Michael B. Jordan	Jamie Foxx	Brie Larson	Drama Crime History	just mercy	Mic B. Joi Je Foxx Lai De
39663	Chinonye Chukwu	Alfre Woodard	Wendell Pierce	Aldis Hodge	Drama	clemency	Woo Wer Pi / Hc Chii

39664 rows × 7 columns

```
In [138]: final_df.isna().sum()
```

```
Out[138]: director_name    0
actor_1_name    0
actor_2_name    0
actor_3_name    0
genres          3
movie_title     0
comb           3
dtype: int64
```

```
In [139]: final_df = final_df.dropna(how='any')
```

```
In [140]: final_df.to_csv('final_data.csv', index=False)
```

```
In [ ]:
```