

Saad Najam

Assignemnt # 3

```
In [26]: import pandas as pd
import numpy as np
```

```
In [2]: data = pd.read_csv("C:\\Users\\User\\Desktop\\python_Assignments\\kaggleFolder\\netflix_titles.csv")
```

```
In [3]: data.head()
```

Out[3]:	show_id	type		title	director	cast	country	date_added	release_year	rating	duration
0	s1	Movie		Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min
1	s2	TV Show		Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Season 1
2	s3	TV Show		Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season
3	s4	TV Show		Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season
4	s5	TV Show		Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Season 1

```
In [4]: data.loc[:, "title"]
```

```
Out[4]: 0      Dick Johnson Is Dead
1      Blood & Water
2      Ganglands
3      Jailbirds New Orleans
4      Kota Factory
...
8802      Zodiac
8803      Zombie Dumb
8804      Zombieland
8805      Zoom
8806      Zubaan
Name: title, Length: 8807, dtype: object
```

```
In [5]: len(data)
```

```
Out[5]: 8807
```

```
In [6]: data.loc[:, ["type"]].value_counts()
```

```
Out[6]: type
Movie      6131
TV Show    2676
dtype: int64
```

```
In [7]: data1 = data.loc[1, "cast"].split(", ")
```

```
In [8]: data1
```

```
Out[8]: ['Ama Qamata',
'Khosi Ngema',
'Gail Mabalane',
'Thabang Molaba',
'Dillon Windvogel',
'Natasha Thahane',
'Arno Greeff',
'Xolile Tshabalala',
'Getmore Sithole',
'Cindy Mahlangu',
'Ryle De Morny',
'Greteli Fincham',
'Sello Maake Ka-Ncube',
'Odwa Gwanya',
'Mekaila Mathys',
'Sandi Schultz',
'Duane Williams',
'Shamilla Miller',
'Patrick Mofokeng']
```

```
In [9]: len(data1)
```

Out[9]: 19

```
In [10]: data2 = ((data.loc[ 1 , "cast"].split(", ")))
```

```
In [11]: print(data2)
```

```
['Ama Qamata', 'Khosi Ngema', 'Gail Mabalane', 'Thabang Molaba', 'Dillon Windvogel', 'Natas  
ha Thahane', 'Arno Greeff', 'Xolile Tshabalala', 'Getmore Sithole', 'Cindy Mahlangu', 'Ryle  
De Morny', 'Greteli Fincham', 'Sello Maake Ka-Ncube', 'Odwa Gwanya', 'Mekaila Mathys', 'San  
di Schultz', 'Duane Williams', 'Shamilla Miller', 'Patrick Mofokeng']
```

```
In [12]: #data['cast'] = data['cast'].fillna(0)  
data.fillna(0)
```

```
Out[12]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	0	United States	September 25, 2021	2020	PG-13	
1	s2	TV Show	Blood & Water	0	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	S
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 :
3	s4	TV Show	Jailbirds New Orleans	0	0	0	September 24, 2021	2021	TV-MA	1 :
4	s5	TV Show	Kota Factory	0	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	S

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	1
8803	s8804	TV Show	Zombie Dumb	0	0	0	July 1, 2019	2018	TV-Y7	S
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	1

8807 rows × 12 columns



Question # 1

In [30]:

```
#data["cast"] = data['cast'].replace(np.nan, "None")
data.fillna(0, inplace=True)

def new_cast(self):

    for i in range( 1 , len(data)):
        return len(data.loc[ i , "cast"].split(", "))
    i = i + 1
```

```
In [31]: data["new_cast"] = data["cast"].apply(new_cast)
```

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

```
Out[32]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
--	---------	------	-------	----------	------	---------	------------	--------------	--------	----------

0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	None	United States	September 25, 2021	2020	PG-13	90 min
---	----	-------	----------------------	-----------------	------	---------------	--------------------	------	-------	--------

1	s2	TV Show	Blood & Water	0	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	1 Season
---	----	---------	---------------	---	---	--------------	--------------------	------	-------	----------

2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season
---	----	---------	-----------	-----------------	---	---	--------------------	------	-------	----------

3	s4	TV Show	Jailbirds New Orleans	0	None	0	September 24, 2021	2021	TV-MA	1 Season
---	----	---------	-----------------------	---	------	---	--------------------	------	-------	----------

4	s5	TV Show	Kota Factory	0	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	1 Season
---	----	---------	--------------	---	---	-------	--------------------	------	-------	----------

ILOC

```
In [33]: data.head()
```

Out[33]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	None	United States	September 25, 2021	2020	PG-13	90 min
1	s2	TV Show	Blood & Water	0	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	1 Season
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	0	September 24, 2021	2021	TV-MA	1 Season
3	s4	TV Show	Jailbirds New Orleans	0	None	0	September 24, 2021	2021	TV-MA	1 Season
4	s5	TV Show	Kota Factory	0	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	1 Season

In [38]:

```
data.loc[data.release_year >= 2020].count()
```

Out[38]:

```
show_id      1545
type         1545
title        1545
director     1545
cast         1545
country      1545
date_added   1545
release_year 1545
rating       1545
duration     1545
listed_in    1545
description   1545
```

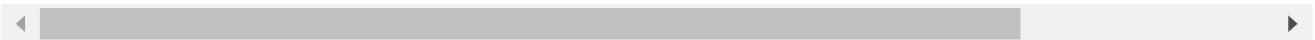
```
new_cast      1545
dtype: int64
```

```
In [41]: data.isnull()
```

```
Out[41]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...
8802	False	False	False	False	False	False	False	False	False	False	False
8803	False	False	False	False	False	False	False	False	False	False	False
8804	False	False	False	False	False	False	False	False	False	False	False
8805	False	False	False	False	False	False	False	False	False	False	False
8806	False	False	False	False	False	False	False	False	False	False	False

8807 rows × 13 columns



```
In [43]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype  
---  -
0   show_id         8807 non-null   object  
1   type            8807 non-null   object  
2   title           8807 non-null   object  
3   director        8807 non-null   object  
4   cast            8807 non-null   object  
5   country         8807 non-null   object  
6   date_added      8807 non-null   object  
7   release_year    8807 non-null   int64   
8   rating          8807 non-null   object  
9   duration        8807 non-null   object  
10  listed_in       8807 non-null   object  
11  description      8807 non-null   object  
12  new_cast        8807 non-null   int64   
dtypes: int64(2), object(11)
memory usage: 894.6+ KB
```

```
In [50]: data.groupby("type")["rating"].value_counts()
```

```
Out[50]: type      rating
Movie      TV-MA      2062
           TV-14      1427
           R          797
           TV-PG      540
           PG-13      490
           PG         287
           TV-Y7      139
           TV-Y       131
           TV-G       126
           NR         75
           G          41
           TV-Y7-FV    5
           NC-17       3
           UR          3
           0           2
           66 min      1
           74 min      1
           84 min      1
TV Show    TV-MA      1145
           TV-14      733
           TV-PG      323
           TV-Y7      195
           TV-Y       176
           TV-G       94
           NR         5
           0           2
           R           2
           TV-Y7-FV    1
Name: rating, dtype: int64
```

```
In [46]: type(data)
```

```
Out[46]: pandas.core.frame.DataFrame
```

Question # 2

```
In [55]: #print(data['type'].value_counts())

#using groupby

data.groupby(['type']).size()
```

```
Out[55]: type
Movie      6131
TV Show    2676
dtype: int64
```

```
In [45]: data.groupby("type").groups
```

```
Out[45]: {'Movie': [0, 6, 7, 9, 12, 13, 16, 18, 22, 23, 24, 26, 27, 28, 29, 30, 35, 36, 38, 41, 42,
43, 44, 45, 46, 48, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 68, 71, 73, 75, 77,
78, 79, 80, 81, 84, 86, 88, 90, 91, 93, 94, 96, 101, 103, 105, 107, 112, 113, 114, 115, 11
6, 117, 118, 119, 122, 126, 127, 128, 129, 130, 131, 133, 134, 135, 136, 137, 138, 139, 14
0, 141, 142, 143, 144, 145, 146, 149, 150, 151, 152, 155, 156, 157, 158, 159, 161, 162,
```