

Netflix DATA Analysis

```
In [1]: import pandas as pd

In [2]: data=pd.read_csv("C:\Users\NMGATA\Downloads\Netflix_data_analysis_project\.netflix_titles.csv")

In [3]: data

Out[3]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0         s1  Movie  Dick Johnson Is Dead  Kirsten Johnson  NaN  United States  September 25, 2021  2020  PG-13  90 min  Documentaries  As her father nears the end of his life, film...
1         1  s2  TV Show  Blood & Water  NaN  Ama Qamata, Khosi Ngema, Gail Mablane, Thabani...  South Africa  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  After crossing paths at a party, a Cape Town l...
2         2  s3  TV Show  Ganglands  Julien Leclercq  Sami Bouajila, Tracy Gotsas, Samuel Jouy, Nabil...  NaN  September 24, 2021  2021  TV-MA  1 Season  Crime TV Shows, International TV Shows, TV Act...  To protect his family from a powerful drug kr...
3         3  s4  TV Show  Jailbirds New Orleans  NaN  NaN  NaN  September 24, 2021  2021  TV-MA  1 Season  Docuseries, Reality TV  Frueh's firations and toilet talk go down amo...
4         4  s5  TV Show  Kota Factory  NaN  Mayur More, Jhendra Kumar, Ranjan Raj, Alam K...  India  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  In a city of coaching centers known to train l...
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
8802  s8802  TV Show  Zombic Dumb  NaN  NaN  NaN  July 1, 2019  2018  TV-R  2 Seasons  Kids' TV Shows, TV Comedies  While living alone in a spooky town, a young g...
8804  s8805  Movie  Zombieland  Ruben Fleischer  Jesse Eisenberg, Woody Harelson, Emma Stone, ...  United States  November 1, 2019  2009  R  88 min  Comedies, Horror Movies  Looking to survive in a world taken over by zo...
8805  s8806  Movie  Zoom  Peter Hewitt  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...  United States  January 11, 2020  2006  PG  88 min  Children & Family Movies, Comedies  Dragged from civilian life, a former superhero...
8806  s8807  Movie  Zubaan  Moeze Singh  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...  India  March 2, 2019  2015  TV-14  111 min  Dramas, International Movies, Music & Musicals  A scrappy but poor boy worms his way into a ty...

8807 rows x 12 columns

In [ ]:
```

Getting Some Basic Information about Dataset

```
In [4]: data.head() # to show first 5 records from dataset

Out[4]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0         s1  Movie  Dick Johnson Is Dead  Kirsten Johnson  NaN  United States  September 25, 2021  2020  PG-13  90 min  Documentaries  As her father nears the end of his life, film...
1         1  s2  TV Show  Blood & Water  NaN  Ama Qamata, Khosi Ngema, Gail Mablane, Thabani...  South Africa  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  After crossing paths at a party, a Cape Town l...
2         2  s3  TV Show  Ganglands  Julien Leclercq  Sami Bouajila, Tracy Gotsas, Samuel Jouy, Nabil...  NaN  September 24, 2021  2021  TV-MA  1 Season  Crime TV Shows, International TV Shows, TV Act...  To protect his family from a powerful drug kr...
3         3  s4  TV Show  Jailbirds New Orleans  NaN  NaN  NaN  September 24, 2021  2021  TV-MA  1 Season  Docuseries, Reality TV  Frueh's firations and toilet talk go down amo...
4         4  s5  TV Show  Kota Factory  NaN  Mayur More, Jhendra Kumar, Ranjan Raj, Alam K...  India  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  In a city of coaching centers known to train l...

In [5]: data.tail() # to show last 5 records from dataset

Out[5]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
8802  s8802  TV Show  Zombic Dumb  NaN  NaN  NaN  July 1, 2019  2018  TV-R  2 Seasons  Kids' TV Shows, TV Comedies  While living alone in a spooky town, a young g...
8804  s8805  Movie  Zombieland  Ruben Fleischer  Jesse Eisenberg, Woody Harelson, Emma Stone, ...  United States  November 1, 2019  2009  R  88 min  Comedies, Horror Movies  Looking to survive in a world taken over by zo...
8805  s8806  Movie  Zoom  Peter Hewitt  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...  United States  January 11, 2020  2006  PG  88 min  Children & Family Movies, Comedies  Dragged from civilian life, a former superhero...
8806  s8807  Movie  Zubaan  Moeze Singh  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...  India  March 2, 2019  2015  TV-14  111 min  Dramas, International Movies, Music & Musicals  A scrappy but poor boy worms his way into a ty...

In [6]: data.shape # this attribute will show the shape of dataframe

Out[6]: (8807, 12)

In [7]: data.size # this attribute will show the total elements present in the dataframe

Out[7]: 105684

In [8]: data.columns # this will show the columns names of the dataset

Out[8]: Index(['show_id', 'title', 'type', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in', 'description'], dtype='object')

In [9]: data.info() #to show information about dataframe like indexes,columns,datatype of each columns, memory size of dataframe

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 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  6173 non-null  object  4  cast  7982 non-null  object  5  country  7976 non-null  object  6  date_added  8787 non-null  object  7  release_year  8807 non-null  int64  8  rating  8803 non-null  object  9  duration  8804 non-null  object  10  listed_in  8807 non-null  object  11  description  8807 non-null  object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB

In [10]: data.dtypes #to return data types of each column in the dataframe

Out[10]:
show_id      object
type          object
title         object
director      object
cast          object
country       object
date_added   object
release_year  int64
rating        object
duration      object
listed_in     object
description   object
dtype: object

In [ ]:
```

Task 1. Is there are any duplicate records?Yes,then removes those records.

```
In [11]: data[data.duplicated()]

Out[11]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description

In [ ]:
```

Task 2. Is there any null values present in dataframe?Show null-values with heat map.

```
In [12]: data.isnull().sum()

Out[12]:
show_id      0
type          0
title         0
director    2634
cast         825
country      831
date_added   10
release_year  0
rating        4
duration      3
listed_in     0
description   0
dtype: object

In [13]: import seaborn as sns # importing seaboar library

In [14]: sns.heatmap(data.isnull())

Out[14]:
<AxesSubplot: >

In [ ]:
```

Distribution of Content:

To begin the task of analyzing Netflix data, I'll start by looking at the distribution of content ratings on Netflix:

```
In [15]: data.head(2)

Out[15]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0         s1  Movie  Dick Johnson Is Dead  Kirsten Johnson  NaN  United States  September 25, 2021  2020  PG-13  90 min  Documentaries  As her father nears the end of his life, film...
1         1  s2  TV Show  Blood & Water  NaN  Ama Qamata, Khosi Ngema, Gail Mablane, Thabani...  South Africa  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  After crossing paths at a party, a Cape Town l...

In [16]: z=data.groupby(['rating']).size().reset_index(name='count')

In [17]: z

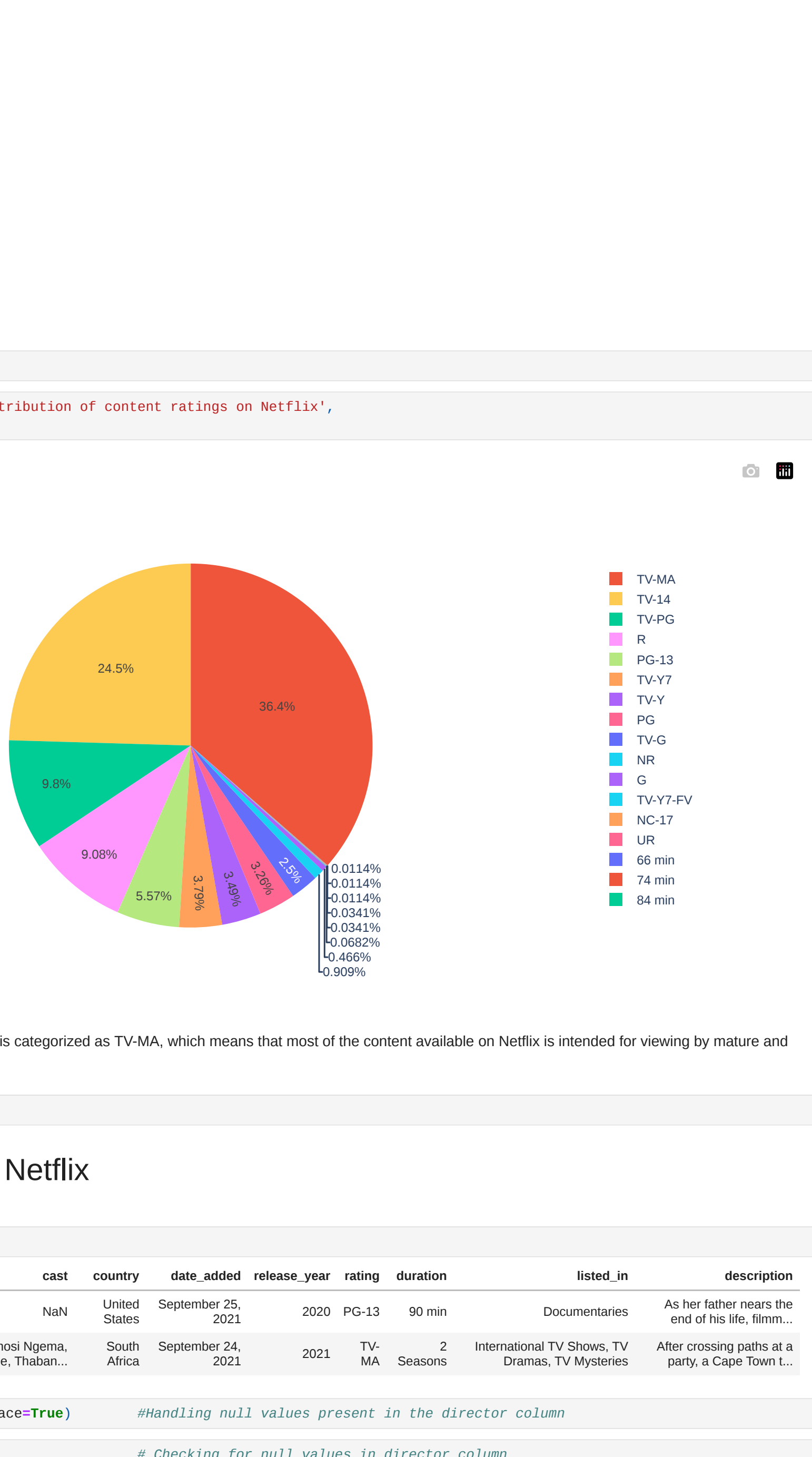
Out[17]:
   rating  count
0    66 min    1
1    74 min    1
2    84 min    1
3         G    41
4   NC-17     3
5      NR     80
6     PG    287
7  PG-13   490
8        R    799
9  TV-14  2190
10     TV-G    220
11  TV-MA  3207
12  TV-PG   863
13     TV-Y   307
14  TV-Y7   334
15  TV-Y7-FV    6
16     UR     31

In [18]: import plotly.express as px

In [19]: px.pie(z,values='count',names='rating',title='Distribution of content ratings on Netflix',color='rating')

In [ ]:
```

Distribution of content ratings on Netflix



The graph above shows that the majority of content on Netflix is categorized as TV-MA, which means that most of the content available on Netflix is intended for viewing by mature and adult audiences.

Top 5 Directors and Actors on Netflix

1.Top 5 Directors on Netflix :

```
In [28]: data.head(2)

Out[28]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0         s1  Movie  Dick Johnson Is Dead  Kirsten Johnson  NaN  United States  September 25, 2021  2020  PG-13  90 min  Documentaries  As her father nears the end of his life, film...
1         1  s2  TV Show  Blood & Water  NaN  Ama Qamata, Khosi Ngema, Gail Mablane, Thabani...  South Africa  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  After crossing paths at a party, a Cape Town l...

In [21]: data.director.fillna('no director specified',inplace=True) #Handling null values present in the director column

In [22]: data.director.isnull().sum() # Checking for null values in director column

Out[22]: 0

In [23]: filtered_directors=pd.DataFrame()

In [24]: filtered_directors=data[data['director'].str.split(' ',expand=True).stack()

Out[25]:
   0  0  Kirsten Johnson
1  0  0  no director specified
2  0  0  Julien Leclercq
3  0  0  no director specified
4  0  0  no director specified
...
8802  0  David Fincher
8803  0  no director specified
8804  0  Ruben Fleischer
8805  0  Peter Hewitt
8806  0  Moeze Singh
Length: 9612, dtype: object

In [26]: filtered_directors=filtered_directors.to_frame()

In [27]: filtered_directors

Out[27]:
   0  0  Kirsten Johnson
1  0  0  no director specified
2  0  0  Julien Leclercq
3  0  0  no director specified
4  0  0  no director specified
...
8802  0  David Fincher
8803  0  no director specified
8804  0  Ruben Fleischer
8805  0  Peter Hewitt
8806  0  Moeze Singh

9612 rows x 1 columns

In [28]: filtered_directors.columns=['Director']

In [29]: filtered_directors

Out[29]:
   Director
0  0  Kirsten Johnson
1  0  0  no director specified
2  0  0  Julien Leclercq
3  0  0  no director specified
4  0  0  no director specified
...
8802  0  David Fincher
8803  0  no director specified
8804  0  Ruben Fleischer
8805  0  Peter Hewitt
8806  0  Moeze Singh

9612 rows x 1 columns

In [30]: directors_top5=filtered_directors.groupby(['Director']).size().reset_index(name='Total Content')

In [31]: directors

Out[31]:
   Director  Total Content
0  Aaron Moorhead         2
1  Aaron Woolf            1
2  Abbas Alibhai Burmawalla 1
3  Abdulhali Al Noor        1
4  Abhinav Shiv Tiwari       1
...  ...  ...
5116  Cagan Imak             1
5117  Isoldi Uggadottir       1
5118  Oskar Thor Axelsson     1
5119  Omer Fank Sorak         2
5120  Senol Sonmez           2

5121 rows x 2 columns

In [32]: directors_top5=directors[directors.Director!='no director specified'].sort_values(by='Total Content',ascending=False).head()

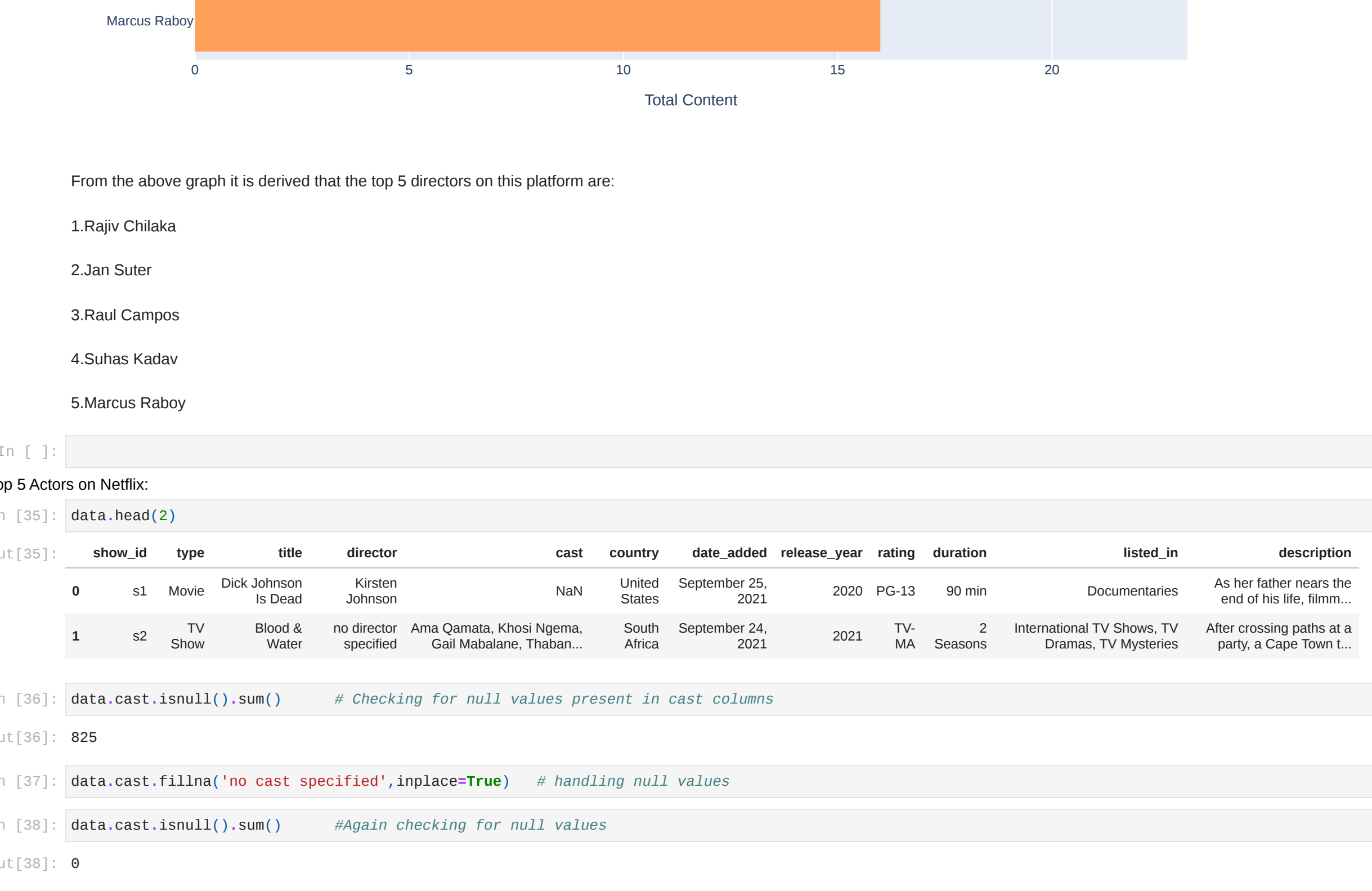
In [33]: directors_top5

Out[33]:
   Director  Total Content
4020  Rajiv Chilaka        22
4067  Raul Campos          18
261  Jan Suter             18
4651  Suhas Kadav          16
3235  Marcus Raboy         16

In [34]: px.bar(directors_top5,x=directors_top5['Total Content'],y=directors_top5['Director'],title='Top 5 Directors on Netflix',color='Director')

In [ ]:
```

Top 5 Directors on Netflix



From the above graph it is derived that the top 5 directors on this platform are:

- 1.Rajiv Chilaka
- 2.Jan Suter
- 3.Raul Campos
- 4.Suhas Kadav
- 5.Marcus Raboy

2.Top 5 Actors on Netflix:

```
In [35]: data.head(2)

Out[35]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0         s1  Movie  Dick Johnson Is Dead  Kirsten Johnson  NaN  United States  September 25, 2021  2020  PG-13  90 min  Documentaries  As her father nears the end of his life, film...
1         1  s2  TV Show  Blood & Water  no director specified  Ama Qamata, Khosi Ngema, Gail Mablane, Thabani...  South Africa  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  After crossing paths at a party, a Cape Town l...

In [36]: data.cast.isnull().sum() # Checking for null values present in cast columns

Out[36]: 825

In [37]: data.cast.fillna('no cast specified',inplace=True) # handling null values

In [38]: data.cast.isnull().sum() #Again checking for null values

Out[38]: 0

In [39]: filtered_cast=pd.DataFrame()

In [40]: filtered_cast=data.cast.str.split(' ',expand=True).stack()

In [41]: filtered_cast

Out[41]:
   0  0  no cast specified
1  0  0  Ama Qamata
2  1  1  Khosi Ngema
3  2  2  Gail Mablane
4  3  3  Thabang Molaba
...
8806  3  Manish Chaudhary
4  4  Meghna Malik
5  5  Malkiet Rauni
6  6  Anita Shabodish
7  7  Chitrangan Tripathy

Length: 64951, dtype: object

In [42]: filtered_cast=filtered_cast.to_frame()

In [43]: filtered_cast

Out[43]:
   0  0  no cast specified
1  0  0  Ama Qamata
2  1  1  Khosi Ngema
3  2  2  Gail Mablane
4  3  3  Thabang Molaba
...
8806  3  Manish Chaudhary
4  4  Meghna Malik
5  5  Malkiet Rauni
6  6  Anita Shabodish
7  7  Chitrangan Tripathy

64951 rows x 1 columns

In [44]: filtered_cast.columns=['Actor']

In [45]: actors=filtered_cast.groupby(['Actor']).size().reset_index(name='Total Content')

In [46]: actors

Out[46]:
   Actor  Total Content
0  Jr.         2
1  "Riley" Lakshar Dridi 1
2  Nalije Dede          1
3  2 Chanz            1
4  2Mex               ...
...  ...  ...
39292  Ibrahim Buiyikal 1
39293  Ibrahim Çelikol 1
39294  Şahin İrmak      1
39295  Şakir Özyıldız   1
39296  Şeyr Dilsu      1

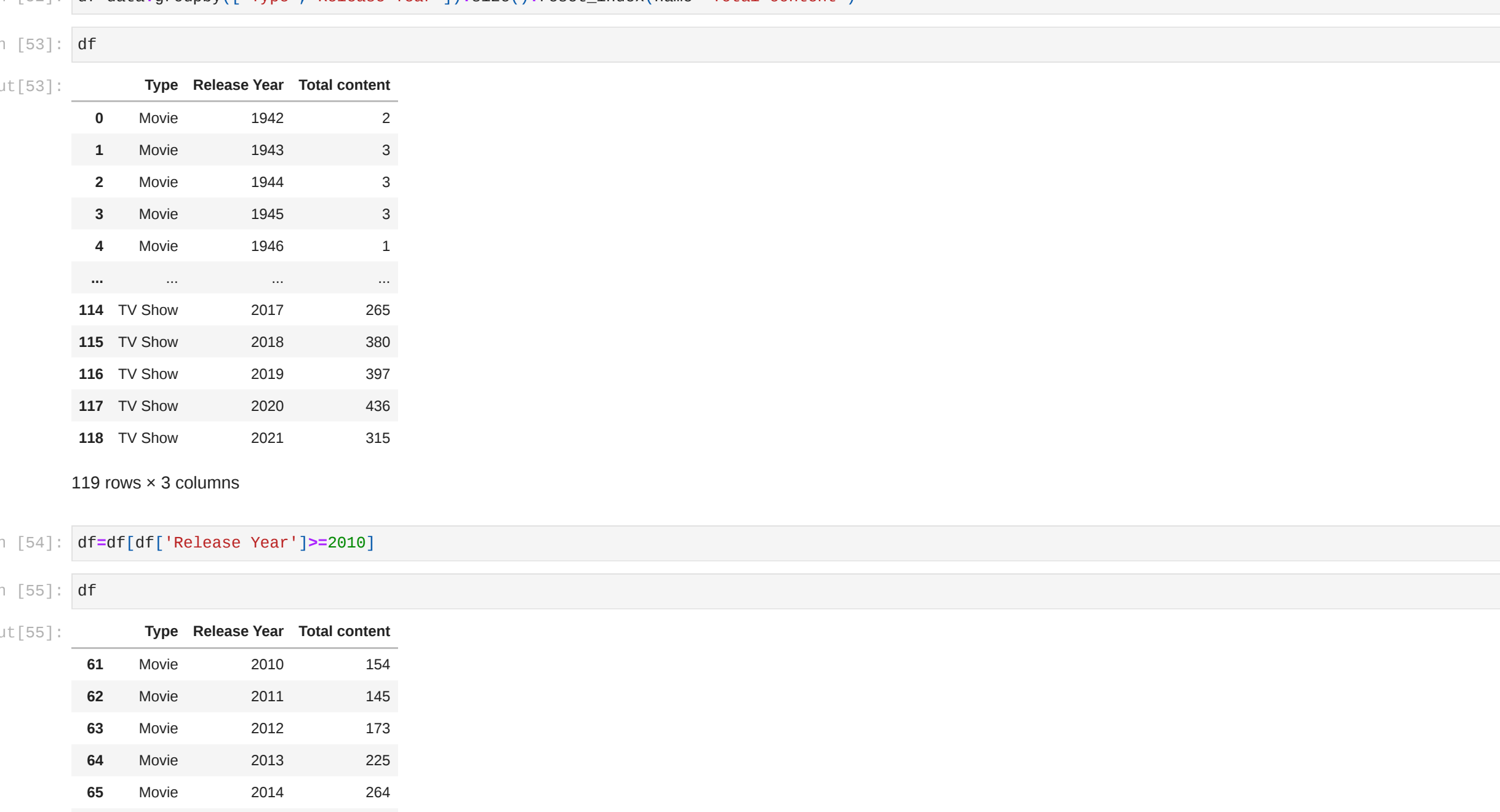
39297 rows x 2 columns

In [47]: actors_top5=actors[actors.Actor!='no cast specified'].sort_values(by='Total Content',ascending=False).head()

In [48]: px.bar(actors_top5,x=actors_top5['Total Content'],y=actors_top5['Actor'],title='Top 5 Actors on Netflix',color='Actor')

In [ ]:
```

Top 5 Actors on Netflix



From the above plot, it is derived that the top 5 actors on Netflix are:

- 1.Anupam Kher
- 2.Rupa Bhatnagar
- 3.Takahiro Sakurai
- 4.Julie Teewani
- 5.Om Puri

The trend of production over the years on Netflix :

```
In [49]: data.head(2)

Out[49]:
   show_id  title  type  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0         s1  Movie  Dick Johnson Is Dead  Kirsten Johnson  NaN  United States  September 25, 2021  2020  PG-13  90 min  Documentaries  As her father nears the end of his life, film...
1         1  s2  TV Show  Blood & Water  no director specified  Ama Qamata, Khosi Ngema, Gail Mablane, Thabani...  South Africa  September 24, 2021  2021  TV-MA  2 Seasons  International TV Shows, TV Dramas, TV Mysteries  After crossing paths at a party, a Cape Town l...

In [50]: data.rename(columns={'release_year':'Release Year','type':'Type'},inplace=True)

In [51]: data[Release Year].unique() # shows the all years present in the data

Out[51]: array([1929, 2021, 1993, 2018, 1996, 1998, 1997, 2019, 2013, 2017, 1975, 1976, 1983, 1987, 2012, 2001, 2014, 2002, 2003, 2004, 2011, 2008, 2009, 2007, 2005, 2006, 1994, 2015, 2019, 2016, 1982, 1989, 1990, 1991, 1999, 1986, 1992, 1984, 1986, 1961, 2009, 1995, 1985, 1976, 1995, 1988, 1981, 1972, 1984, 1945, 1954, 1979, 1950, 1956, 1963, 1970, 1973, 1925, 1974, 1960, 1966, 1971, 1962, 1969, 1977, 1967, 1965, 1946, 1946, 1942, 1955, 1944, 1947, 1943], dtype=int64)

In [52]: df=data.groupby(['Type','Release Year']).size().reset_index(name='Total Content')

In [53]: df

Out[53]:
   Type  Release Year  Total Content
0  Movie  1942             2
1  Movie  1945             3
2  Movie  1946             3
3  Movie  1945             3
4  Movie  1946             1

...

114  TV Show  2017             265
115  TV Show  2018             380
116  TV Show  2019             397
117  TV Show  2020             436
118  TV Show  2021             315

119 rows x 3 columns

In [54]: df[df['df['Release Year']>=2010]

In [55]: df

Out[55]:
   Type  Release Year  Total Content
61  Movie  2010             154
62  Movie  2011             145
63  Movie  2012             173
64  Movie  2013             225
65  Movie  2014             264
66  Movie  2015             358
67  Movie  2016             698
68  Movie  2017             767
69  Movie  2018             633
70  Movie  2019             617
71  Movie  2020             510
72  Movie  2021             470
107  TV Show  2010             40
108  TV Show  2011             64
109  TV Show  2012             64
110  TV Show  2013             63
111  TV Show  2014             88
112  TV Show  2015             162
113  TV Show  2016             244
114  TV Show  2017             265
115  TV Show  2018             380
116  TV Show  2019             397
117  TV Show  2020             436
118  TV Show  2021             315

In [56]: px.line(df,x='Release Year',y='Total Content',title='Trend of content produced over the years on Netflix',color='Type')

In [ ]:
```

Trend of content produced over the years on Netflix



The above line graph shows that there has been a decline in the production of the content for movies since 2018 and for TV Shows since 2020

```
In [ ]:
```