In [1]:
```
import pandas as pd
import numpy as np
from pandas_profiling import ProfileReport
```

```
C:\Users\shivesh\AppData\Local\Temp\ipykernel_2152\2887495799.py:3: DeprecationWarnin
g: `import pandas_profiling` is going to be deprecated by April 1st. Please use `impo
rt ydata_profiling` instead.
  from pandas_profiling import ProfileReport
```

In [2]:
```
netflix = pd.read_csv('netflix.csv')
```

In [3]:
```
netflix.head(5)
```

Out[3]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Docur
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Inte TV S
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dr H
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	C Famil (
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Ind Inte

In [4]:
```
pd.isnull(netflix).sum().sum()
```

Out[4]:
```
0
```

In [5]:
```
netflix.value_counts('type')
```

Out[5]:
```
type
Movie      6126
TV Show    2664
dtype: int64
```

In [6]:
```
netflix.shape
```

Out[6]:
```
(8790, 10)
```

In [7]:
```
netflix.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64 
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

In [8]:
```
netflix['duration'].max()
```

Out[8]:
```
'99 min'
```

In [9]:
```
netflix.sort_values('director',ascending=False).head(5)
```

Out[9]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	
1008	s983	Movie	Hayat Öpücügü	Şenol Sönmez	Turkey	4/23/2021	2015	TV-14	98 min	C Inte
2685	s3534	Movie	Kill Me If You Dare	Şenol Sönmez	Turkey	9/6/2019	2019	TV-14	100 min	C Inte
4350	s6071	Movie	Aşk Tesadüfleri Sever	Ömer Faruk Sorak	Turkey	4/4/2018	2011	TV-MA	122 min	Inte
1007	s982	Movie	G.O.R.A	Ömer Faruk Sorak	Turkey	4/23/2021	2004	TV-MA	123 min	A C Inte
4267	s5976	Movie	Ég man þig	Óskar Thór Axelsson	Iceland	3/27/2018	2017	TV-MA	106 min	Inte

In [10]:
```
netflix.sort_values('director',ascending=True).head(5)
```

Out[10]:

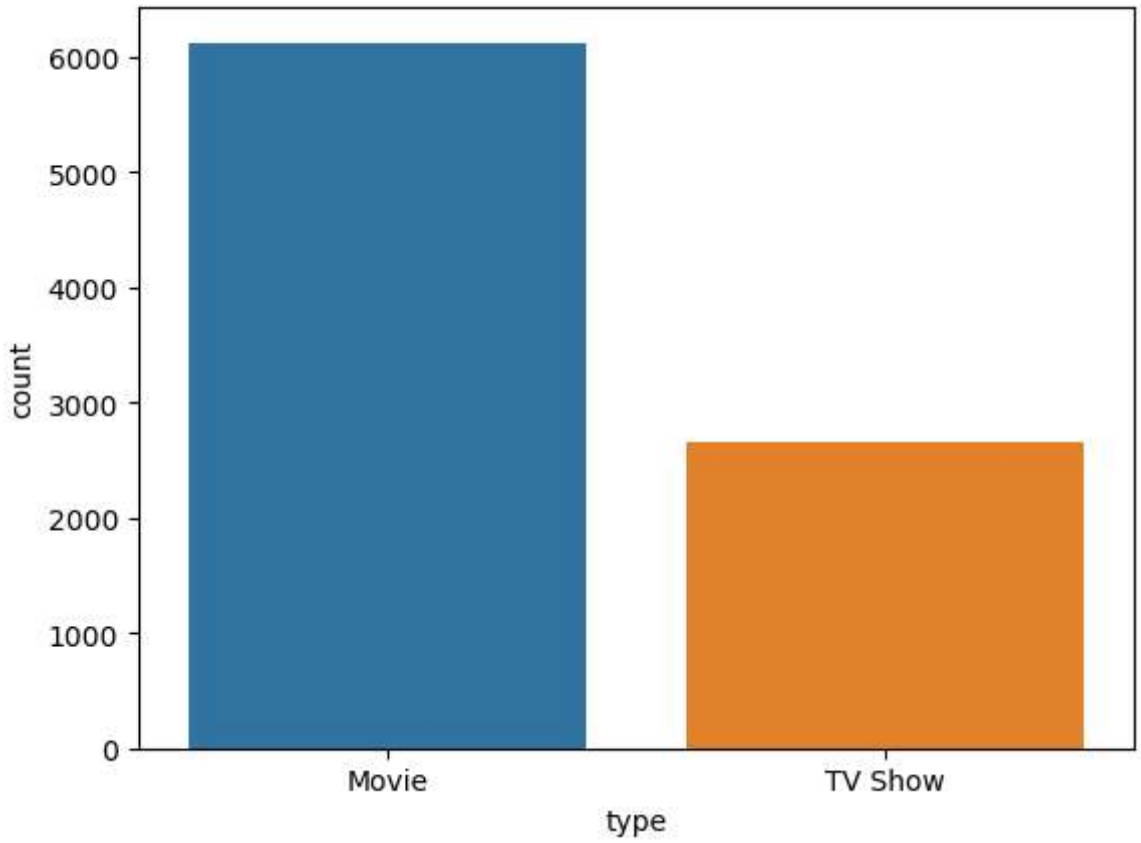
	show_id	type	title	director	country	date_added	release_year	rating	duration
2688	s3538	Movie	Watchman	A. L. Vijay	India	9/4/2019	2019	TV-14	93 min
4357	s6079	Movie	Abhinetri	A. L. Vijay	India	5/1/2018	2016	TV-14	131 min
1952	s2390	Movie	Asura Guru	A. Raajdheep	India	6/13/2020	2020	TV-14	117 min
3977	s5550	Movie	Salaakhen	A. Salaam	India	4/1/2017	1975	TV-14	134 min
2980	s4050	Movie	Sarkar	A.R. Murugadoss	India	3/2/2019	2018	TV-MA	162 min

In [11]:
```
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [12]:
```
sns.countplot(netflix['type'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning:
Pass the following variable as a keyword arg: x. From version 0.12, the only valid po
sitional argument will be `data`, and passing other arguments without an explicit key
word will result in an error or misinterpretation.
  warnings.warn(
```

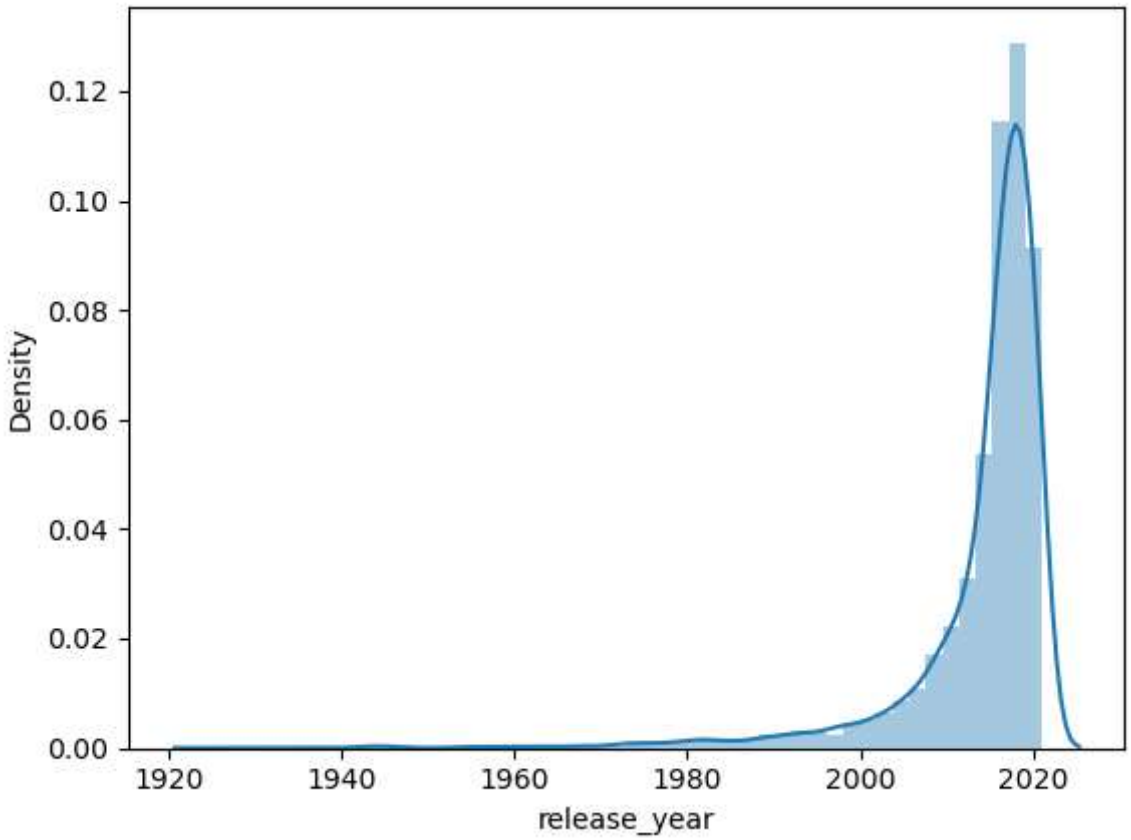
Out[12]:
```
<AxesSubplot:xlabel='type', ylabel='count'>
```



In [7]:
```
sns.distplot(netflix['release_year'])
```

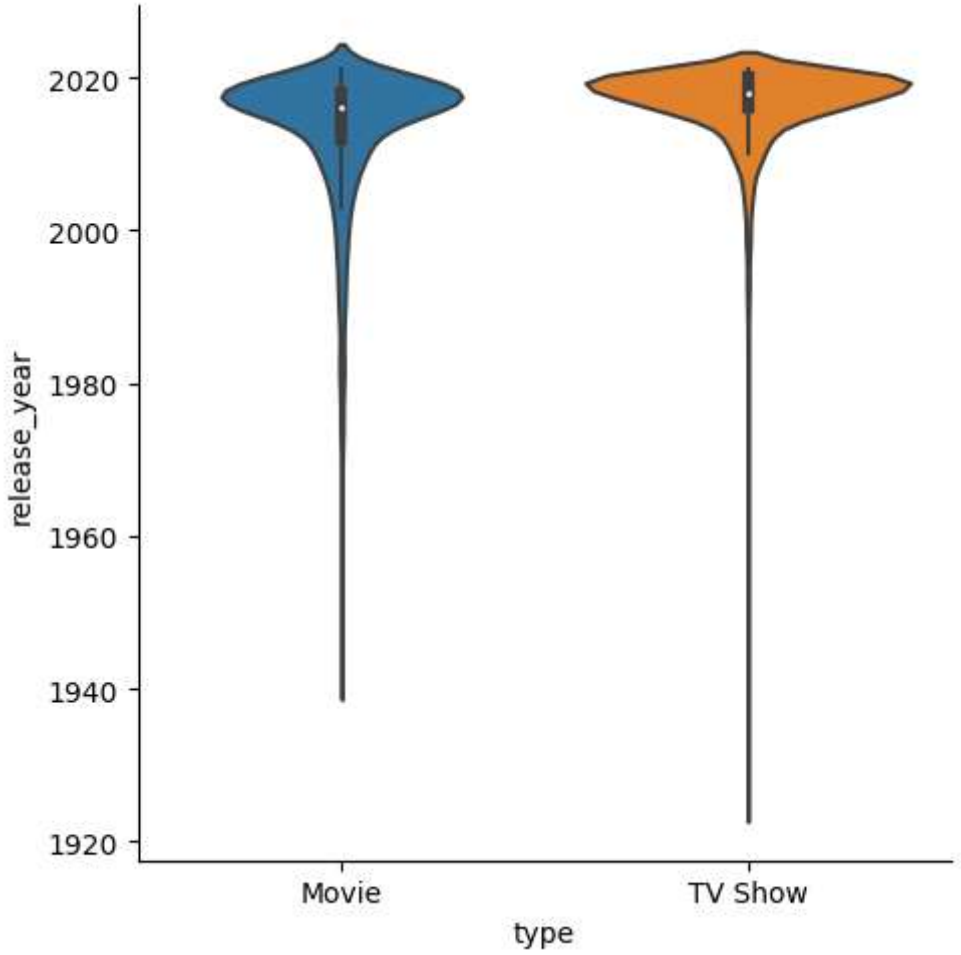
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarni
ng: `distplot` is a deprecated function and will be removed in a future version. Plea
se adapt your code to use either `displot` (a figure-level function with similar flex
ibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[7]:
```
<AxesSubplot:xlabel='release_year', ylabel='Density'>
```



In [8]:
```
sns.catplot(x='type',y='release_year',data=netflix,kind="violin")
```

Out[8]:
```
<seaborn.axisgrid.FacetGrid at 0x223f8908c40>
```



In [26]:
```
profile = ProfileReport(netflix)
```

In [27]:
```
profile
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]
Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]
```

Overview

Dataset statistics

Number of variables	10
Number of observations	8790
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	686.8 KiB
Average record size in memory	80.0 B

Variable types

Categorical	9
Numeric	1

Alerts

`show_id` has a high cardinality: 8790 distinct values	High cardinality
`title` has a high cardinality: 8787 distinct values	High cardinality
`director` has a high cardinality: 4528 distinct values	High cardinality

Out[27]:

In [29]:
```
from autoviz.AutoViz_Class import AutoViz_Class
```

```
Imported v0.1.58. After importing, execute '%matplotlib inline' to display charts in
Jupyter.
    AV = AutoViz_Class()
    dfte = AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0, verbose=1, l
owess=False,
                  chart_format='svg',max_rows_analyzed=150000,max_cols_analyzed=30, save
_plot_dir=None)
Update: verbose=0 displays charts in your local Jupyter notebook.
        verbose=1 additionally provides EDA data cleaning suggestions. It also displa
ys charts.
        verbose=2 does not display charts but saves them in AutoViz_Plots folder in l
ocal machine.
        chart_format='bokeh' displays charts in your local Jupyter notebook.
        chart_format='server' displays charts in your browser: one tab for each chart
type
        chart_format='html' silently saves interactive HTML files in your local machi
ne
```

In [30]:
```
AV=AutoViz_Class()
```

In [ ]:
```
%matplotlib inline
netflix = AV.AutoViz('netflix.csv')
```

```
Shape of your Data Set loaded: (8790, 10)
############################################################################################
##
######################### C L A S S I F Y I N G  V A R I A B L E S  ######################
##
############################################################################################
##
Classifying variables in data set...
Data cleaning improvement suggestions. Complete them before proceeding to ML modelin
g.
```

	Nuniques	dtype	Nulls	Nullpercent	NuniquePercent	Value counts Min	Data cleaning improvement suggestions
show_id	8790	object	0	0.000000	100.000000	1	combine rare categories, possible ID column: drop
title	8787	object	0	0.000000	99.965870	1	combine rare categories
director	4528	object	0	0.000000	51.513083	1	combine rare categories
date_added	1713	object	0	0.000000	19.488055	1	combine rare categories
listed_in	513	object	0	0.000000	5.836177	1	combine rare categories
duration	220	object	0	0.000000	2.502844	1	combine rare categories
country	86	object	0	0.000000	0.978385	1	combine rare categories
release_year	74	int64	0	0.000000	0.841866	0	
rating	14	object	0	0.000000	0.159272	3	
type	2	object	0	0.000000	0.022753	2664	

```
    10 Predictors classified...
Cannot visualize this dataset since no numeric or integer vars in data...returning
'str' object has no attribute 'shape'
Exception Drawing Scatter Plots
Could not draw some Scatter Plots
Could not draw some Distribution Plots
Could not draw some Violin Plots
Could not draw some Heat Maps
Could not draw some Time Series plots
Could not draw some Pivot Charts against Dependent Variable
Could not draw some Bar Charts
```

```
Traceback (most recent call last):
  File "C:\Users\shivesh\AppData\Roaming\Python\Python39\site-packages\autoviz\AutoViz_Class.py", line 429, in AutoViz_Main
    svg_data = draw_scatters(dft,continuous_vars,
  File "C:\Users\shivesh\AppData\Roaming\Python\Python39\site-packages\autoviz\AutoViz_Utils.py", line 284, in draw_scatters
    if dfin.shape[0] >= 10000 or lowess == False:
AttributeError: 'str' object has no attribute 'shape'
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     C:\Users\shivesh\AppData\Roaming\nltk_data...
```

In [ ]: