

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) ○

File + % Run C Code

Library Call

```
In [1]: 1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 import plotly.express as px
```

Data Preparation

```
In [2]: 1 nf=pd.read_csv(r"C:\Users\yuvraj\Downloads\netflix.csv",encoding='Latin')
```

```
In [3]: 1 nf.head()
```

```
Out[3]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	...	Unnamed: 16	Unnamed: 17	Unnamed: 18	Unnamed: 19
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	...	NaN	NaN	NaN	NaN
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	...	NaN	NaN	NaN	NaN
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	...	NaN	NaN	NaN	NaN
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	...	NaN	NaN	NaN	NaN
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alami K...	India	September 24, 2021	2021	TV-MA	2 Seasons	...	NaN	NaN	NaN	NaN

5 rows × 26 columns

```
In [4]: 1 nf.columns
```

```
Out[4]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', '...', 'listed_in', 'description'], dtype='object')
```

Resize the Dataset

```
In [5]: 1 nf=nf.iloc[:,12:]
```

```
In [6]: 1 nf.head()
```

```
Out[6]:
```

	show_id	type	title	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	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	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 t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	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 lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	...	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alami K...	India	September 24, 2021	2021	TV-MA	2 Seasons	...	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

```
In [7]: 1 # Basic Information
```

```
In [8]: 1 nf.shape
```

```
Out[8]: (8809, 12)
```

```
In [9]: 1 nf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8809 entries, 0 to 8808
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   show_id     8809 non-null   object 
 1   type        8809 non-null   object 
 2   title       8809 non-null   object 
 3   director    8809 non-null   object 
 4   cast        8809 non-null   object 
 5   country     8809 non-null   object 
 6   date_added  8809 non-null   datetime64[ns]
 7   release_year 8809 non-null   int64  
 8   rating      8809 non-null   float64
 9   duration    8809 non-null   int64  
 10  listed_in   8809 non-null   object 
 11  description  8809 non-null   object
```

```

2   title      8899 non-null  object
3 director     6175 non-null  object
4   cast       7984 non-null  object
5 country      7978 non-null  object
6 date_added   8799 non-null  object
7 release_year 8809 non-null  int64
8   rating     8805 non-null  object
9   duration    8806 non-null  object
10 listed_in    8809 non-null  object
11 description  8809 non-null  object
dtypes: int64(1), object(11)
memory usage: 826.0+ KB

```

```
In [10]: 1 nf.isna().sum()
```

```

Out[10]: 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: int64

```

```
In [11]: 1 #Filling NaN Values
```

```

2
3 nf.director = nf.director.fillna("No Name")
4 nf.cast    = nf.cast.fillna("No Cast")
5 nf.country = nf.country.fillna("No Country")
6 nf.date_added= nf.date_added.fillna("No Date")
7 nf.rating  = nf.rating.fillna("No Rating")
8 nf.duration= nf.duration.fillna("No Duration")

```

```
In [12]: 1 nf.isna().sum()
```

```

Out[12]: show_id      0
          type        0
          title       0
          director    0
          cast        0
          country     0
          date_added   0
          release_year 0
          rating       0
          duration     0
          listed_in    0
          description  0
          dtype: int64

```

```
In [13]: 1 nf.head()
```

```
Out[13]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No Cast	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmmaker...
1	s2	TV Show	Blood & Water	No Name	Ama Qamata, Khosi Ngema, Gail Mabalane, 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 t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabil...	No Country	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lord...
3	s4	TV Show	Jailbirds New Orleans	No Name	No Cast	No Country	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	No Name	Mayur More, Jitendra Kumar, Ranjan Raj, Alami K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

Groupby(Type)

```

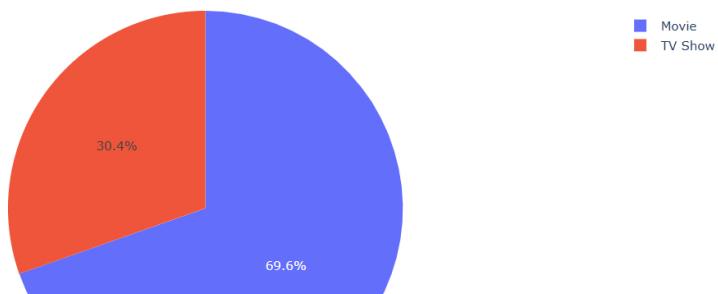
In [14]: 1 movie_tv=nf.groupby(['type']).size().reset_index(name='count')
2 movie_tv

```

```
Out[14]:
```

type	count
Movie	6132
TV Show	2677

```
In [15]: 1 ply.pie(movie_tv,names='type',values='count')
```



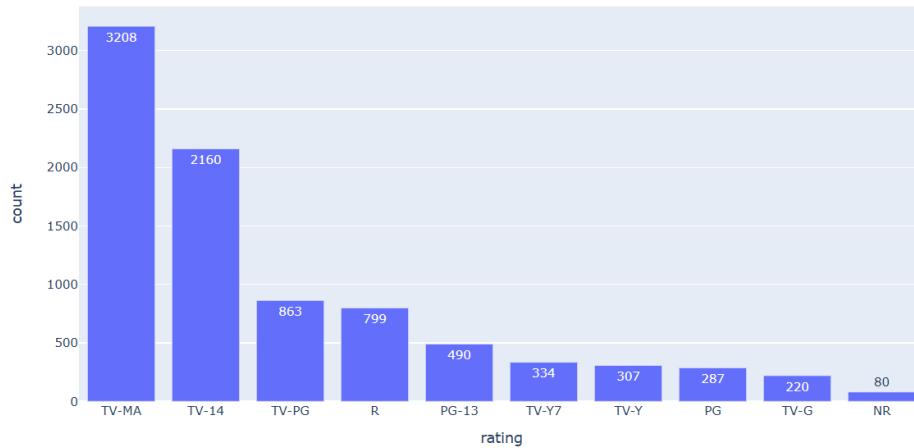
Content Analysis

```
In [16]: 1 content=nf.groupby(by=['rating']).size().reset_index(name='count').nlargest(10,'count')
2 content.head()
```

Out[16]:

	rating	count
13	TV-MA	3208
11	TV-14	2160
14	TV-PG	863
10	R	799
9	PG-13	490

```
In [17]: 1 ply.bar(content,x='rating',y='count',text_auto=True)
```



Top Director

```
In [18]: 1 #stack diffrent name that belong to the same cell
2
3 director = nf['director'].str.split(',',expand=True).stack()
4 director = pd.DataFrame(director,columns=['Director'])
5 director = nf.groupby(by=['director']).size().reset_index(name='count').nlargest(15,'count')
6 director
```

Out[18]:

	director	count
3051	No Name	2634
3393	Rajiv Chilaka	19
3444	RaÃºl Campos, Jan Suter	18
2598	Marcus Raboy	16
4047	Suhas Kadav	16
1790	Jay Karas	14
685	Cathy Garcia-Molina	13
1787	Jay Chapman	12
2671	Martin Scorsese	12
4482	Youssef Chahine	12
4021	Steven Spielberg	11
1105	Don Michael Paul	10
973	David Dhawan	9
1282	Fernando AyllÃ³n	8
1607	Hakan AlgÃ¼l	8

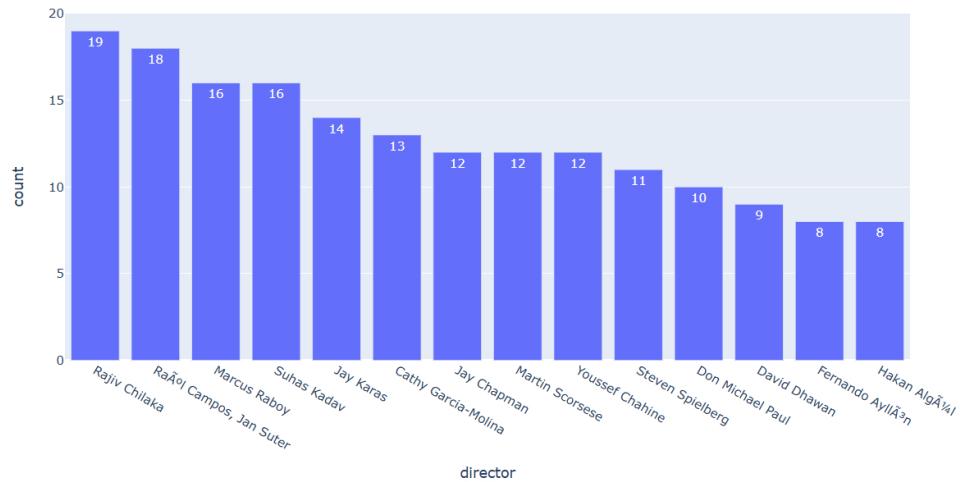
```
In [19]: 1 director=director[director['director'] !='No Name']
2 director
```

Out[19]:

	director	count
3393	Rajiv Chilaka	19
3444	RaÃºl Campos, Jan Suter	18
2598	Marcus Raboy	16
4047	Suhas Kadav	16
1790	Jay Karas	14
685	Cathy Garcia-Molina	13
4707	Jay Chapman	12

ID	Cast Member	Count
2671	Martin Scorsese	12
4482	Youssef Chahine	12
4021	Steven Spielberg	11
1105	Don Michael Paul	10
973	David Dhawan	9
1282	Fernando Ayllón	8
1507	Hakan Algül	8

```
In [20]: 1 ply.bar(director,x='director',y='count',text_auto=True)
```



Top Cast

```
In [21]: 1 #stack individual
2 casting_individual=nmf['cast'].str.split(",").expand=True).stack().reset_index(drop=True)
3 casting_individual
```

```
Out[21]: 0           No Cast
1           Ama Qamata
2           Khosi Ngema
3           Gail Mabalane
4           Thabang Molaba
...
64985       Jiří Maria Sieber
64986       Raymond Waring
64987       Petr Drozda
64988       John Comer
64989       Benedetta Degli Innocenti
Length: 64990, dtype: object
```

```
In [22]: 1 cast_individual=pd.DataFrame(casting_individual,columns=['TotalCast'])
2 cast_individual.head()
```

```
Out[22]:
TotalCast
0    No Cast
1    Ama Qamata
2    Khosi Ngema
3    Gail Mabalane
4    Thabang Molaba
```

```
In [23]: 1 top_15_cast= cast_individual.groupby(by=['TotalCast']).size().reset_index(name='count')
2 top_15_cast= top_15_cast.sort_values(by=['count'],ascending=False)
3 top_15_cast.head()
```

```
Out[23]:
TotalCast  count
37675     No Cast   825
2612      Anupam Kher   39
26965     Rupa Bhimani   31
30327     Takahiro Sakurai   30
15555     Julie Tejwani   28
```

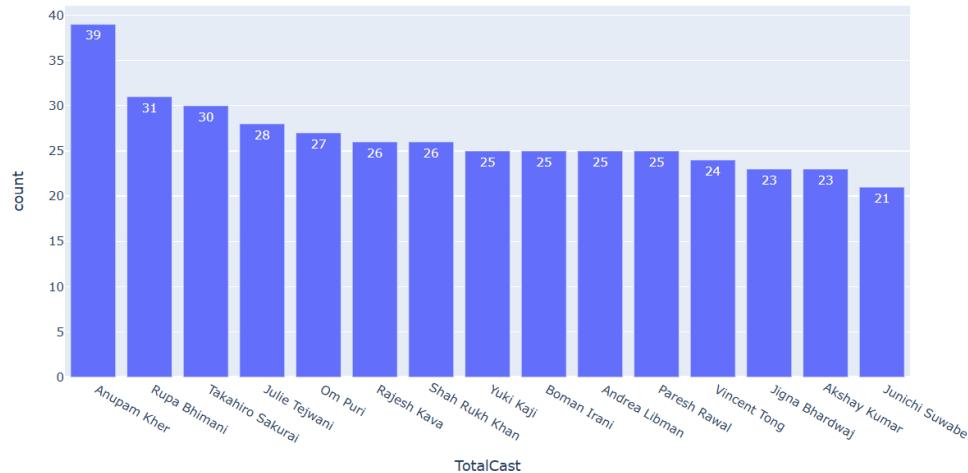
```
In [24]: 1 top_15_cast= top_15_cast.iloc[1:,:]
```

```
In [25]: 1 top_15_cast= top_15_cast.nlargest(15,'count')
```

```
In [26]: 1 top_15_cast.head()
```

```
Out[26]:
TotalCast  count
2612     Anupam Kher   39
26965    Rupa Bhimani   31
30327    Takahiro Sakurai   30
15555    Julie Tejwani   28
23642    Om Puri      27
```

```
In [27]: 1 plt.bar(top_15_cast.x='TotalCast',y='count',text_auto=True)
```



Listed In

```
In [28]: 1 #Stack diffrent name that belong to same cell
2 new_listed_in = nf['Listed_in'].str.split(',',expand=True).stack().reset_index(drop=True)
3 new_listed_in = pd.DataFrame(new_listed_in,columns=['Listed'])
4 new_listed_in.head()
```

Out[28]:

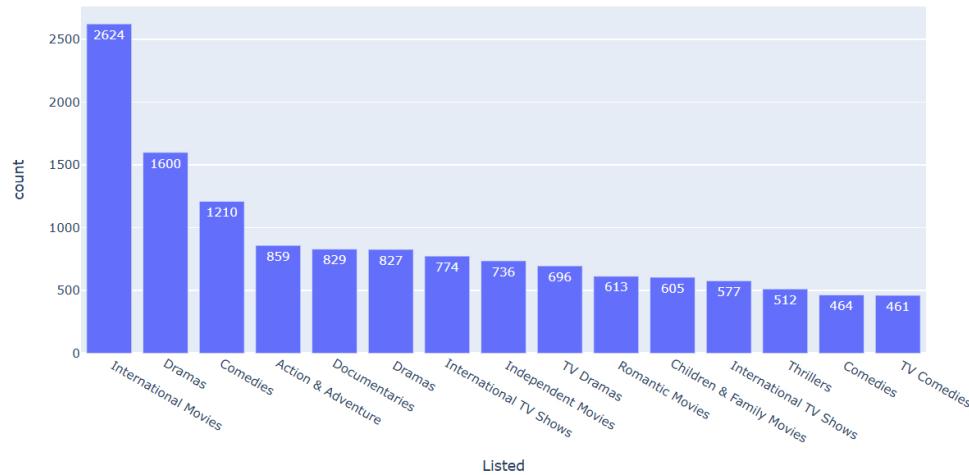
	Listed
0	Documentaries
1	International TV Shows
2	TV Dramas
3	TV Mysteries
4	Crime TV Shows

```
In [29]: 1 top_15_item = new_listed_in.groupby(by=['Listed']).size().reset_index(name='count')
2 top_15_item = top_15_item.nlargest(15,'count')
3 top_15_item
```

Out[29]:

	Listed	count
15	International Movies	2624
54	Dramas	1600
48	Comedies	1210
41	Action & Adventure	859
51	Documentaries	829
10	Dramas	827
58	International TV Shows	774
14	Independent Movies	736
33	TV Dramas	696
23	Romantic Movies	613
45	Children & Family Movies	605
16	International TV Shows	577
40	Thrillers	512
5	Comedies	464
32	TV Comedies	461

```
In [30]: 1 plt.bar(top_15_item,x='Listed',y='count',text_auto=True)
```



Type vs Country

```
In [31]: 1 content=nf.groupby(by=["type", 'country']).size().reset_index(name='count').nlargest(16,'count')
2 content.head()
```

```
Out[31]:
```

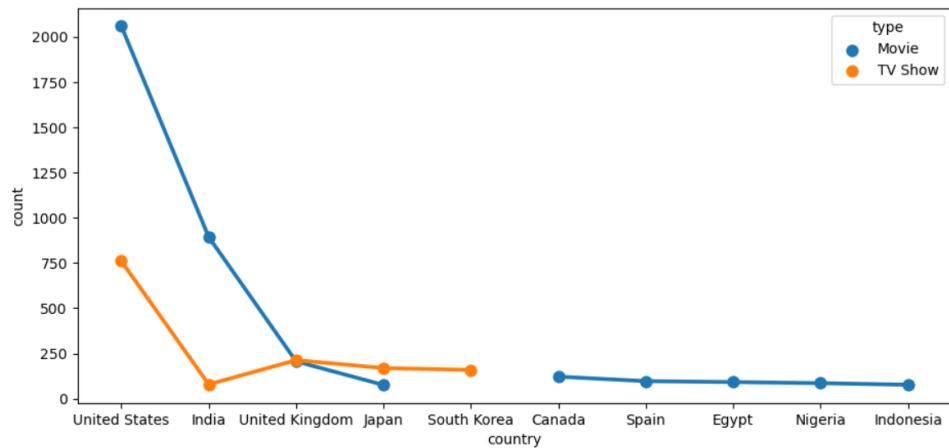
	type	country	count
526	Movie	United States	2059
218	Movie	India	893
813	TV Show	United States	760
321	Movie	No Country	440
754	TV Show	No Country	391

```
In [32]: 1 content=content[content['country'] != 'No Country']
2 content
```

```
Out[32]:
```

	type	country	count
526	Movie	United States	2059
218	Movie	India	893
813	TV Show	United States	760
793	TV Show	United Kingdom	213
441	Movie	United Kingdom	206
735	TV Show	Japan	169
773	TV Show	South Korea	159
50	Movie	Canada	122
385	Movie	Spain	97
128	Movie	Egypt	92
319	Movie	Nigeria	86
718	TV Show	India	79
238	Movie	Indonesia	77
278	Movie	Japan	76

```
In [33]: 1 plt.figure(figsize=(11,5))
2 sns.pointplot(x='country',y='count',hue='type',data=content)
3 plt.show()
```



Type and Release Year

```
In [34]: 1 type_and_year = nf.groupby(by=['type','release_year']).size().reset_index(name='count')
2 type_and_year = type_and_year[type_and_year['release_year']>2007]
3 type_and_year.head()
```

```
Out[34]:
```

	type	release_year	count
59	Movie	2008	113
60	Movie	2009	118
61	Movie	2010	154
62	Movie	2011	145
63	Movie	2012	173

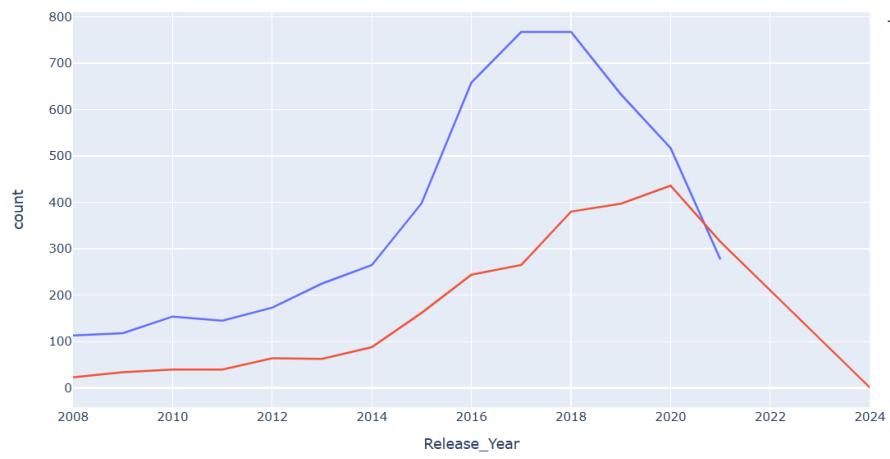
```
In [35]: 1 type_and_year.rename(columns={'type':'Type','release_year':'Release_Year'},inplace=True)
```

```
In [36]: 1 type_and_year.head()
```

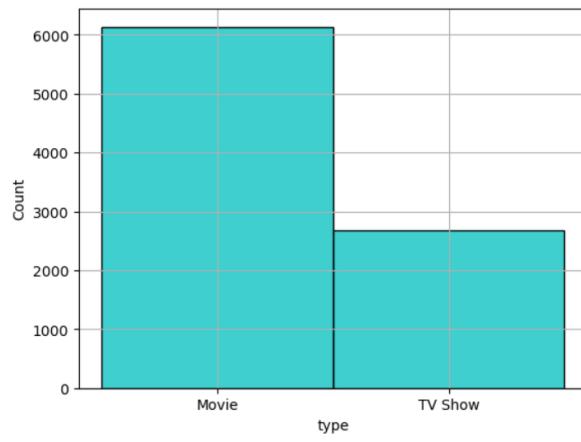
```
Out[36]:
```

	Type	Release_Year	count
59	Movie	2008	113
60	Movie	2009	118
61	Movie	2010	154
62	Movie	2011	145
63	Movie	2012	173

```
In [37]: 1 plt.line(type_and_year,x="Release_Year",y="count",color='Type')
```



```
In [38]: 1 sns.histplot(x="type",data=nf,edgecolor="black",bins=8,color="c")
2 plt.grid()
```



```
In [ ]: 1
```

```
In [ ]: 1
```

```
In [ ]: 1
```