

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
import datetime as dt
import warnings
warnings.filterwarnings('ignore')
```

```
In [3]: data=pd.read_csv("C:/Users/0ooba/Desktop/Analysis with python/NETFLIX/netflix.csv")
data
```

Out[3]:

| | show_id | type | | title | director | country | date_added | release_year | rating | duration | listed_in |
|------|---------|---------|--|----------------------------------|-----------------|---------------|------------|--------------|--------|-----------|---------------------------------------------------|
| 0 | s1 | Movie | | Dick Johnson Is Dead | Kirsten Johnson | United States | 9/25/2021 | 2020 | PG-13 | 90 min | Documentaries |
| 1 | s3 | TV Show | | Ganglands | Julien Leclercq | France | 9/24/2021 | 2021 | TV-MA | 1 Season | Crime TV Shows, International TV Shows, TV Act... |
| 2 | s6 | TV Show | | Midnight Mass | Mike Flanagan | United States | 9/24/2021 | 2021 | TV-MA | 1 Season | TV Dramas, TV Horror, TV Mysteries |
| 3 | s14 | Movie | | Confessions of an Invisible Girl | Bruno Garotti | Brazil | 9/22/2021 | 2021 | TV-PG | 91 min | Children & Family Movies, Comedies |
| 4 | s8 | Movie | | Sankofa | Haile Gerima | United States | 9/24/2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies |
| ... | ... | ... | | ... | ... | ... | ... | ... | ... | ... | ... |
| 8785 | s8797 | TV Show | | Yunus Emre | Not Given | Turkey | 1/17/2017 | 2016 | TV-PG | 2 Seasons | International TV Shows, TV Dramas |
| 8786 | s8798 | TV Show | | Zak Storm | Not Given | United States | 9/13/2018 | 2016 | TV-Y7 | 3 Seasons | Kids' TV |
| 8787 | s8801 | TV Show | | Zindagi Gulzar Hai | Not Given | Pakistan | 12/15/2016 | 2012 | TV-PG | 1 Season | International TV Shows, Romantic TV Shows, TV ... |
| 8788 | s8784 | TV Show | | Yoko | Not Given | Pakistan | 6/23/2018 | 2016 | TV-Y | 1 Season | Kids' TV |
| 8789 | s8786 | TV Show | | YOM | Not Given | Pakistan | 6/7/2018 | 2016 | TV-Y7 | 1 Season | Kids' TV |

8790 rows × 10 columns

Data cleaning

```
In [4]: data.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 [5]: data['date_added'] = pd.to_datetime(data['date_added'], errors='coerce', infer_datetime_format=True)
data.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   datetime64[ns]
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: datetime64[ns](1), int64(1), object(8)
memory usage: 686.8+ KB

```

```
In [6]: data.describe()
```

```

Out[6]:
release_year
count  8790.000000
mean   2014.183163
std     8.825466
min    1925.000000
25%    2013.000000
50%    2017.000000
75%    2019.000000
max    2021.000000

```

```
In [7]: data.drop("show_id", axis=1)
```

```

Out[7]:
   type  title  director  country  date_added  release_year  rating  duration  listed_in
0  Movie  Dick Johnson Is Dead  Kirsten Johnson  United States  2021-09-25  2020  PG-13  90 min  Documentaries
1  TV Show  Ganglands  Julien Leclercq  France  2021-09-24  2021  TV-MA  1 Season  Crime TV Shows, International TV Shows, TV Act...
2  TV Show  Midnight Mass  Mike Flanagan  United States  2021-09-24  2021  TV-MA  1 Season  TV Dramas, TV Horror, TV Mysteries
3  Movie  Confessions of an Invisible Girl  Bruno Garotti  Brazil  2021-09-22  2021  TV-PG  91 min  Children & Family Movies, Comedies
4  Movie  Sankofa  Haile Gerima  United States  2021-09-24  1993  TV-MA  125 min  Dramas, Independent Movies, International Movies
...  ...  ...  ...  ...  ...  ...  ...  ...  ...
8785  TV Show  Yunus Emre  Not Given  Turkey  2017-01-17  2016  TV-PG  2 Seasons  International TV Shows, TV Dramas
8786  TV Show  Zak Storm  Not Given  United States  2018-09-13  2016  TV-Y7  3 Seasons  Kids' TV
8787  TV Show  Zindagi Gulzar Hai  Not Given  Pakistan  2016-12-15  2012  TV-PG  1 Season  International TV Shows, Romantic TV Shows, TV ...
8788  TV Show  Yoko  Not Given  Pakistan  2018-06-23  2016  TV-Y  1 Season  Kids' TV
8789  TV Show  YOM  Not Given  Pakistan  2018-06-07  2016  TV-Y7  1 Season  Kids' TV

```

8790 rows × 9 columns

```
In [8]: data.isna().sum()
```

```

Out[8]:
show_id    0
type       0
title      0
director   0
country    0
date_added 0
release_year 0
rating     0
duration   0
listed_in  0
dtype: int64

```

```
In [9]: data.duplicated().sum()
```

```

Out[9]:
0

```

Data Manipulation

```
In [11]: types =data['type'].value_counts()  
types
```

```
Out[11]: Movie      6126  
TV Show    2664  
Name: type, dtype: int64
```

```
In [12]: Top_country=data["country"].value_counts().head(10)  
Top_country
```

```
Out[12]: United States    3240  
India                    1057  
United Kingdom          638  
Pakistan                 421  
Not Given                287  
Canada                   271  
Japan                    259  
South Korea              214  
France                   213  
Spain                    182  
Name: country, dtype: int64
```

```
In [13]: Top_director_movie= data["director"][data["type"]=="Movie"].value_counts().sort_values(ascending=False).head(7)  
Top_director_movie.columns=["director","Movie count"]  
Top_director_movie
```

```
Out[13]:
```

| | director | Movie count |
|---|------------------------|-------------|
| 0 | Not Given | 173 |
| 1 | Rajiv Chilaka | 19 |
| 2 | Raúl Campos, Jan Suter | 18 |
| 3 | Suhas Kadav | 16 |
| 4 | Marcus Raboy | 15 |
| 5 | Jay Karas | 14 |
| 6 | Cathy Garcia-Molina | 13 |

```
In [14]: Top_director_TV= data["director"][data["type"]=="TV Show"].value_counts().sort_values(ascending=False).head(7).  
Top_director_TV.columns=["director","TV count"]  
Top_director_TV
```

```
Out[14]:
```

| | director | TV count |
|---|------------------------------|----------|
| 0 | Not Given | 2415 |
| 1 | Alastair Fothergill | 14 |
| 2 | Mark Thornton, Todd Kauffman | 5 |
| 3 | Stan Lathan | 3 |
| 4 | Hsu Fu-chun | 2 |
| 5 | Iginio Straffi | 2 |
| 6 | Shin Won-ho | 2 |

```
In [15]: Top_rating_Movie =data["rating"][data["type"]=="Movie"].value_counts().sort_values(ascending=False).head(10).re  
Top_rating_Movie.columns=["rating","Movie count"]  
Top_rating_Movie
```

```
Out[15]:
```

| | rating | Movie count |
|---|--------|-------------|
| 0 | TV-MA | 2062 |
| 1 | TV-14 | 1427 |
| 2 | R | 797 |
| 3 | TV-PG | 540 |
| 4 | PG-13 | 490 |
| 5 | PG | 287 |
| 6 | TV-Y7 | 139 |
| 7 | TV-Y | 131 |
| 8 | TV-G | 126 |
| 9 | NR | 75 |

```
In [16]: Top_rating_TV= data['rating'][data['type'] == 'TV Show'].value_counts().sort_values(ascending=False).head(10).r  
Top_rating_TV.columns=["rating","TV count"]  
Top_rating_TV
```

```
Out[16]:
```

| | rating | TV count |
|---|----------|----------|
| 0 | TV-MA | 1143 |
| 1 | TV-14 | 730 |
| 2 | TV-PG | 321 |
| 3 | TV-Y7 | 194 |
| 4 | TV-Y | 175 |
| 5 | TV-G | 94 |
| 6 | NR | 4 |
| 7 | R | 2 |
| 8 | TV-Y7-FV | 1 |

```
In [17]: data.rename(columns={'listed_in' : 'category'}, inplace=True)
```

```
In [18]: Top_category=data["category"].value_counts().head(10)
Top_category
```

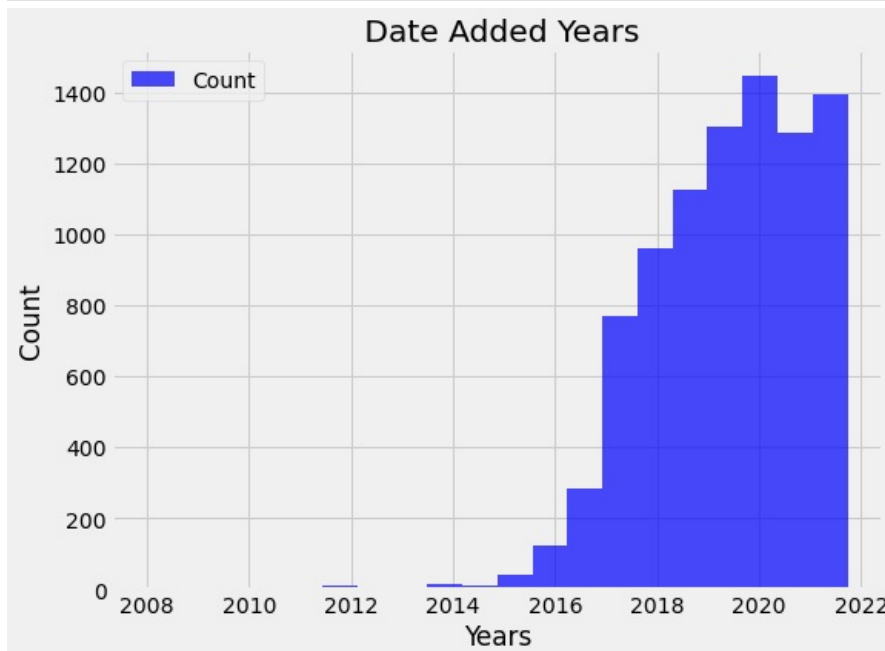
```
Out[18]:
```

| | |
|--------------------------------------------------|-----|
| Dramas, International Movies | 362 |
| Documentaries | 359 |
| Stand-Up Comedy | 334 |
| Comedies, Dramas, International Movies | 274 |
| Dramas, Independent Movies, International Movies | 252 |
| Kids' TV | 219 |
| Children & Family Movies | 215 |
| Children & Family Movies, Comedies | 201 |
| Documentaries, International Movies | 186 |
| Dramas, International Movies, Romantic Movies | 180 |

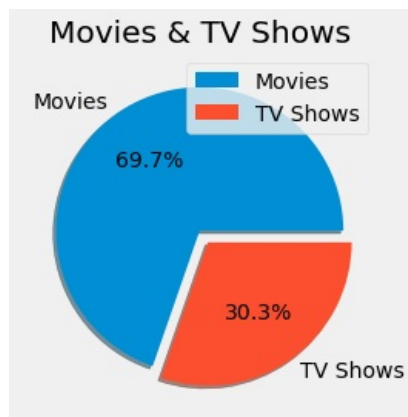
Name: category, dtype: int64

Data Visualization

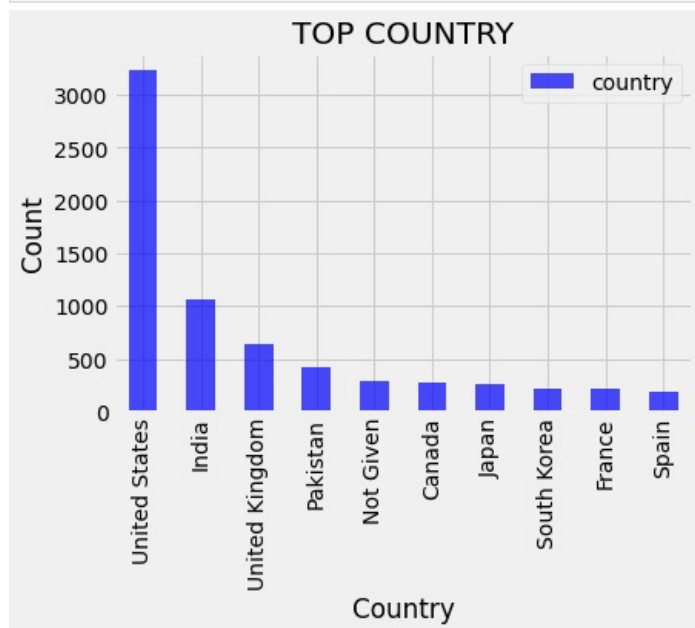
```
In [29]: plt.figure(figsize=(8,6))
plt.style.use("fivethirtyeight")
plt.hist(data['date_added'], bins=20, label='Count', color = "b",alpha=0.7 );
plt.grid(True)
plt.axis("equal")
plt.title(' Date Added Years')
plt.xlabel('Years')
plt.ylabel('Count')
plt.legend()
plt.show()
```



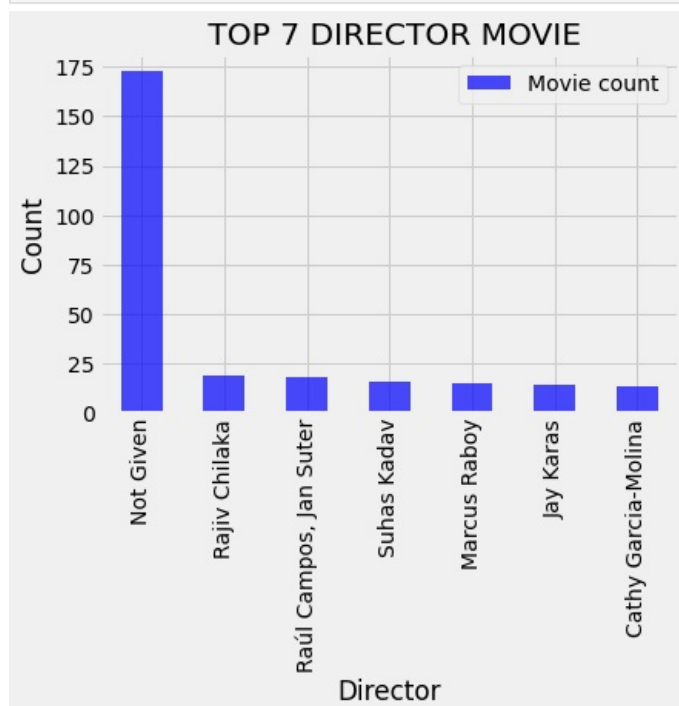
```
In [27]: plt.pie(types ,labels=['Movies', 'TV Shows'], autopct='%1.1f%%', explode=[0,0.1],shadow=True)
plt.title('Movies & TV Shows')
plt.legend()
plt.show()
```



```
In [34]: Top_country.plot(kind="bar",color="b",alpha=0.7)
plt.title("TOP COUNTRY")
plt.xlabel("Country")
plt.ylabel("Count")
plt.legend()
plt.show()
```

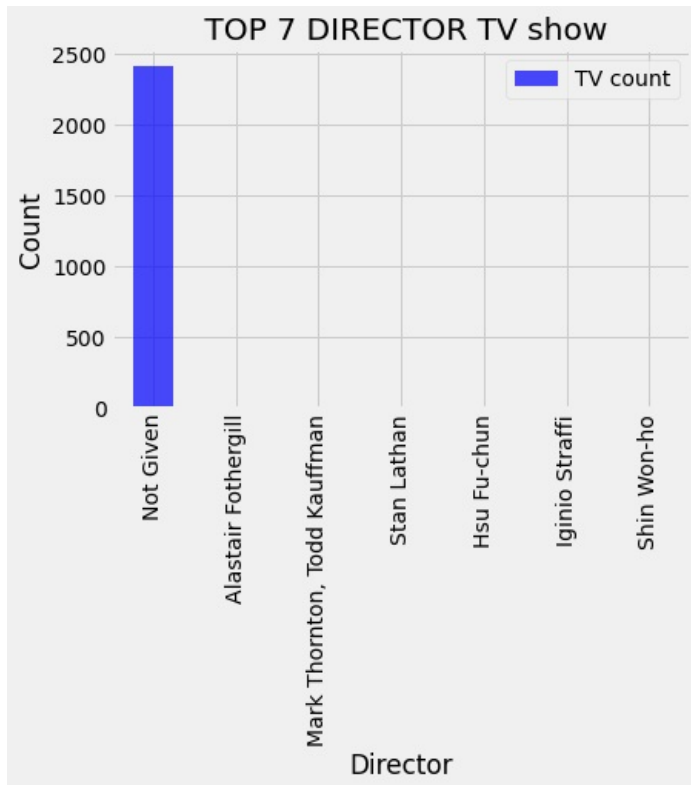


```
In [35]: Top_director_movie.plot(kind="bar",x="director",y="Movie count",color="b",alpha=0.7)
plt.title(" TOP 7 DIRECTOR MOVIE")
plt.xlabel("Director")
plt.ylabel("Count")
plt.legend()
plt.show()
```

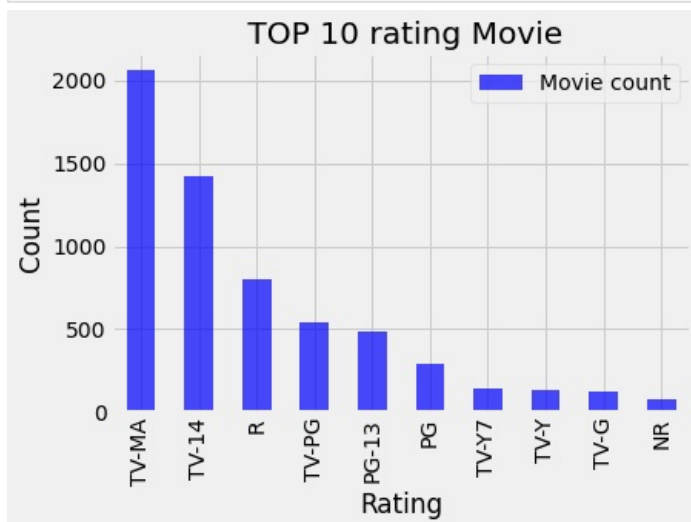


```
In [36]: Top_director_TV.plot(kind="bar",x="director",y="TV count",color="b",alpha=0.7)
```

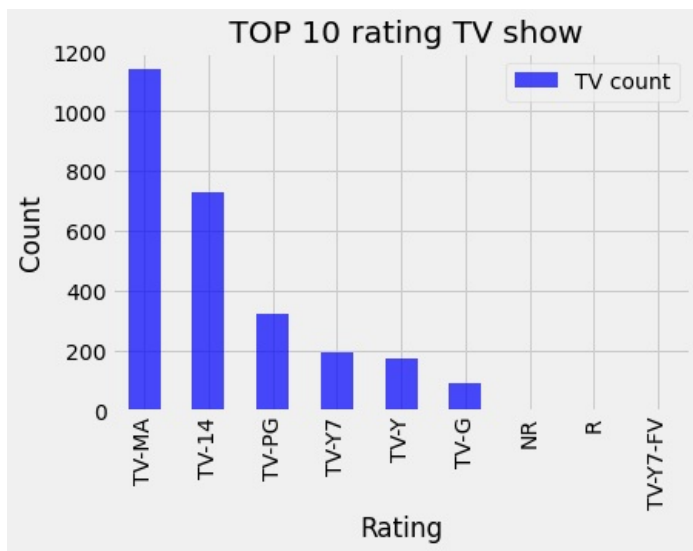
```
plt.title(" TOP 7 DIRECTOR TV show")
plt.xlabel("Director")
plt.ylabel("Count")
plt.legend()
plt.show()
```



```
In [37]: Top_rating_Movie.plot(kind="bar",x="rating",y="Movie count",color="b",alpha=0.7)
plt.title(" TOP 10 rating Movie")
plt.xlabel("Rating")
plt.ylabel("Count")
plt.legend()
plt.show()
```



```
In [38]: Top_rating_TV.plot(kind="bar",x="rating",y="TV count",color="b",alpha=0.7)
plt.title(" TOP 10 rating TV show")
plt.xlabel("Rating")
plt.ylabel("Count")
plt.legend()
plt.show()
```



```
In [39]: Top_category.plot(kind="bar",color="b",alpha=0.7)
plt.title(" TOP 10 category")
plt.xlabel("Categoryr")
plt.ylabel("Count")
plt.legend()
plt.show()
```

