

EXPLORATORY DATA ANALYSIS ON NETFLIX

BY
KRISHNAMURTHI G

Table of Content

- About this Project.
- Problem statements.
- Dataset
- Solutions

NETFLIX

About this Project

➤ Details of Dataset.

- In this project we have used a Dataset of Netflix in which it contain the Movies and TV shows, by using this dataset we have performed the content analysis.
- This dataset contains 11 columns and 7789 rows.

➤ Programming Language used:


- Python

1. Packages Name

1. Pandas
2. Numpy
3. Matplotlib
4. Seaborn

Problem Statements

- Checking duplicate records in dataset.
- Removing the duplicate records.
- Checking null values present in any column.
- Using heap-map showing null values count.
- Checking how many movies and TV shows are in the dataset? Using bar graph.

- 
- A large, semi-transparent red watermark of the Netflix logo is centered in the background of the slide.
- **Checking Category Content based on Country.**
 - **Checking major ratings given to Netflix Shows.**
 - **Most of the shows has TV-MA and TV-14 ratings.**
 - **Finding Popular Cast Analysis.**
 - **In which year the highest number of TV shows and Movies where Released? Showing with Bar Graph.**

Dataset

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	Horror Movies, International Movies	When an army recruit is found dead, his fellow...
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...	In a postapocalyptic world, rag-doll robots hi...
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas	A brilliant group of students become card-coun...
5	s6	TV Show	46	Serdar Akar	Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan...	Turkey	July 1, 2017	TV-MA	1 Season	International TV Shows, TV Dramas, TV Mysteries	A genetics professor experiments with a treatm...
6	s7	Movie	122	Yasir Al Yasiri	Amina Khalil, Ahmed Dawood, Tarek Lotfy, Ahmed...	Egypt	June 1, 2020	TV-MA	95 min	Horror Movies, International Movies	After an awful accident, a couple admitted to ...
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	November 1, 2019	R	119 min	Dramas	After one of his high school students attacks ...
8	s9	Movie	706	Shravan Kumar	Divya Dutta, Atul Kulkarni, Mohan Agashe, Anup...	India	April 1, 2019	TV-14	118 min	Horror Movies, International Movies	When a doctor goes missing, his psychiatrist w...
9	s10	Movie	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	December 15, 2017	TV-MA	143 min	Horror Movies, International Movies, Thrillers	An architect and his wife move into a castle t...

Solutions

Checking duplicate records in dataset

```
In [12]: data.shape
```

```
Out[12]: (7789, 11)
```

```
In [13]: #To Duplicate  
data[data.duplicated()]           #to check wheter the data(table) contains duplcate values are not
```

```
Out[13]:
```

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
6300	s684	Movie	Backfire	Dave Patten	Black Deniro, Byron "Squally" Vinson, Dominic ...	United States	April 5, 2019	TV-MA	97 min	Dramas, Independent Movies, Thrillers	When two would-be robbers accidentally kill a ...
6622	s6621	Movie	The Lost Okoroshi	Abba T. Makama	Seun Ajayi, Judith Audu, Tope Tedela, Ifu Enna...	Nigeria	September 4, 2020	TV-MA	94 min	Comedies, Dramas, Independent Movies	A disillusioned security guard transforms into...

Removing the duplicate records

To remove the duplicate record

```
In [14]: # To Delete Duplicates          # To Remove the duplicate values
data.drop_duplicates(inplace=True)      #without 'inplace' the data will not store permanetlly with inplace=true data will store per
```

```
In [15]: data[data.duplicated()] #Duplicates are removed
```

Out[15]:

Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
---------	----------	-------	----------	------	---------	--------------	--------	----------	------	-------------

```
In [16]: data.shape
```

Out[16]: (7787, 11)

Checking null values present in any column

```
In [18]: data.isnull() #Shows wheter the data(table) has null values are not
```

Out[18]:

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
0	False	False	False	True	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
...
7784	False	False	False	False	False	False	False	False	False	False	False
7785	False	False	False	False	False	False	False	False	False	False	False
7786	False	False	False	True	False	True	False	False	False	False	False
7787	False	False	False	True	False	False	False	False	False	False	False
7788	False	False	False	False	True	False	False	False	False	False	False

7787 rows × 11 columns

```
In [19]: data.isnull().sum() # To get the count of Null values in each Column
```

Out[19]:

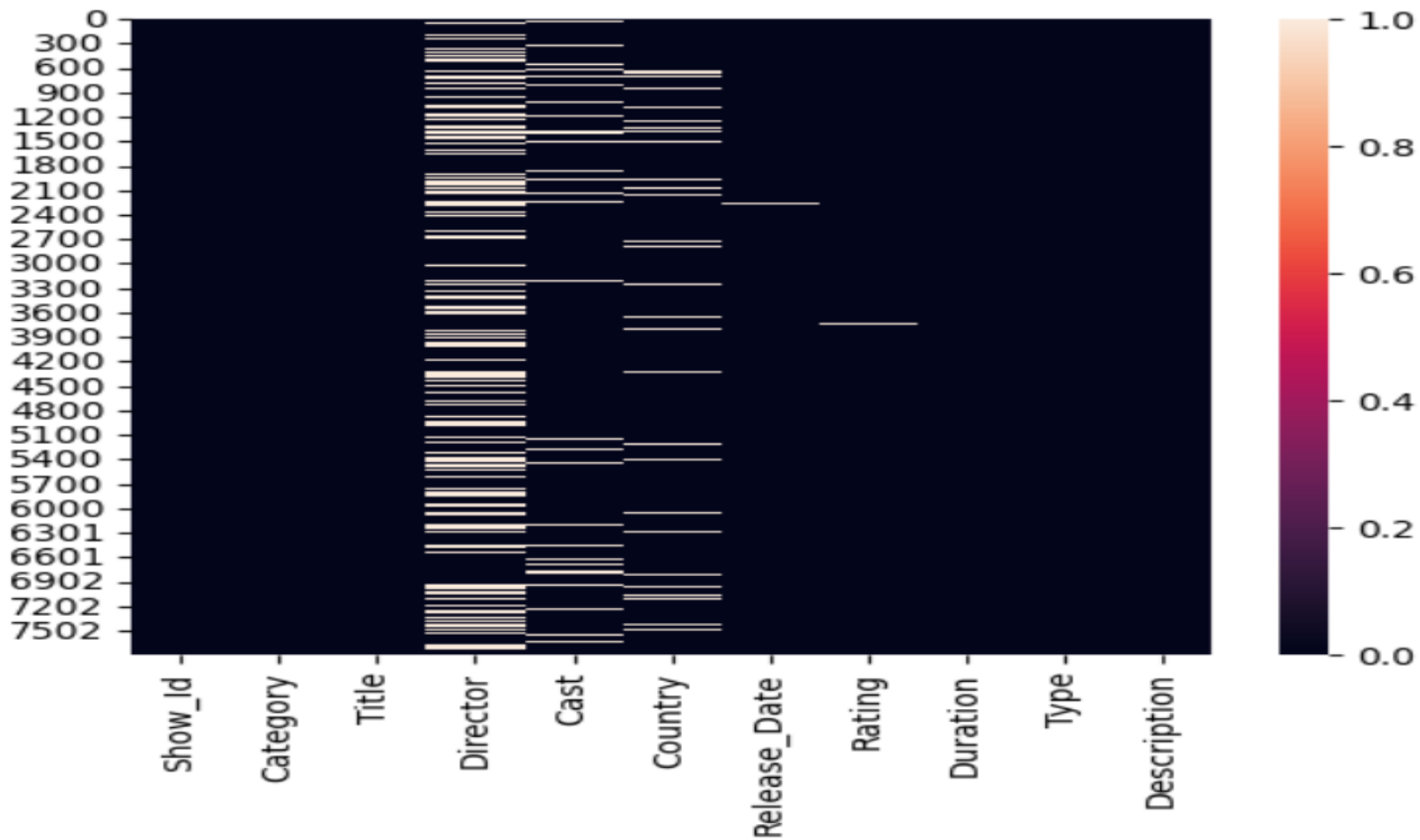
Show_Id	0
Category	0
Title	0
Director	2388
Cast	718
Country	507
Release Date	10

Using heap-map showing null values count

```
In [23]: sns.heatmap(data.isnull())
```

#Using heap-map to show null values count

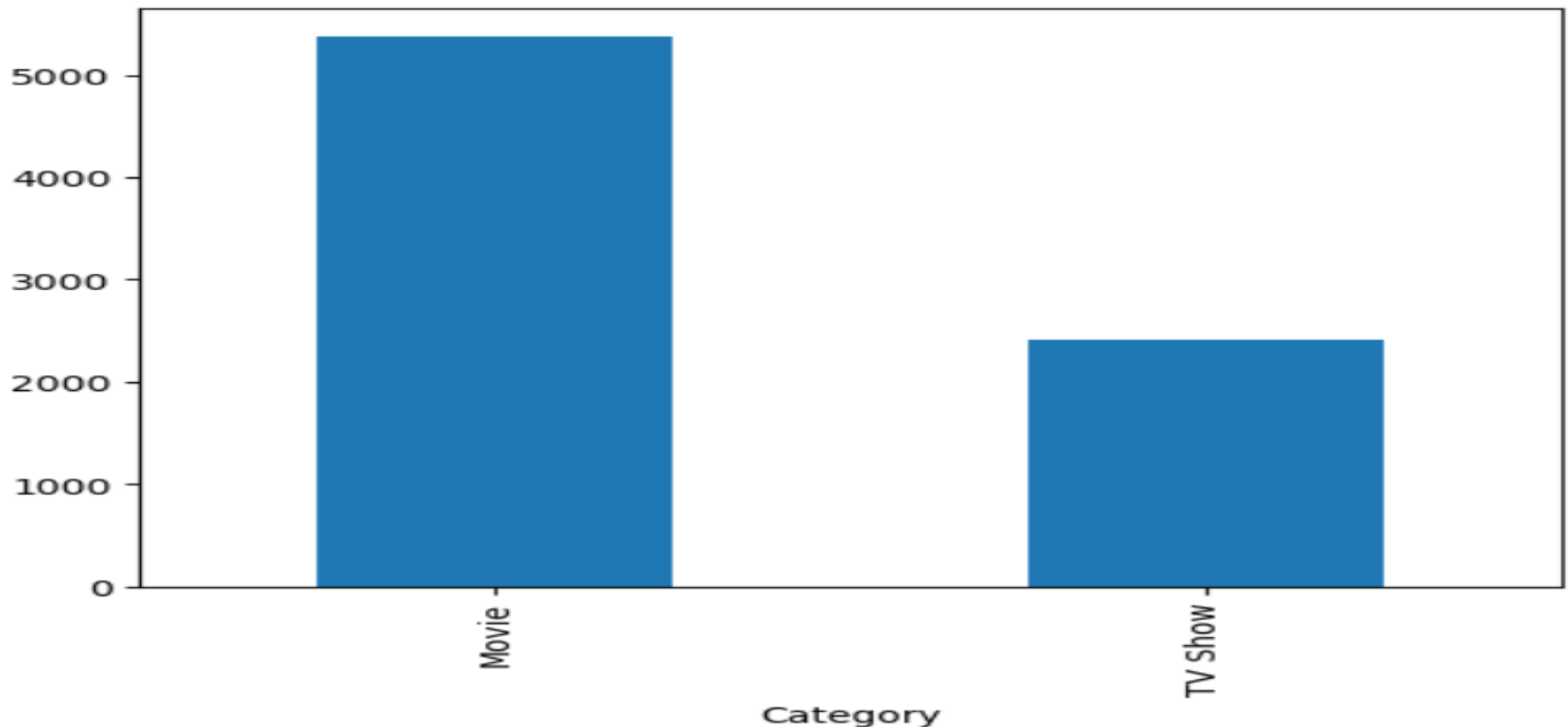
```
Out[23]: <AxesSubplot:>
```



Checking how many movies and TV shows are in the dataset? Using bar graph.

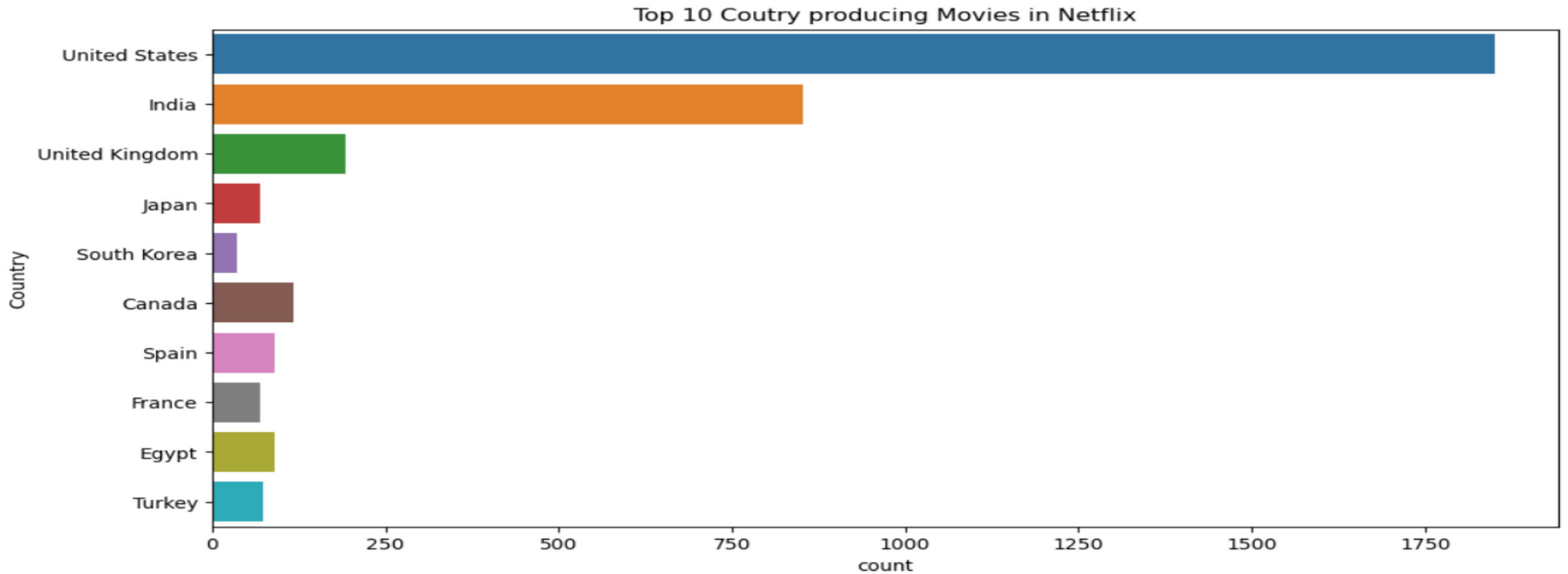
```
In [52]: data.groupby("Category").Category.count().plot(kind='bar')
```

```
Out[52]: <AxesSubplot:xlabel='Category'>
```



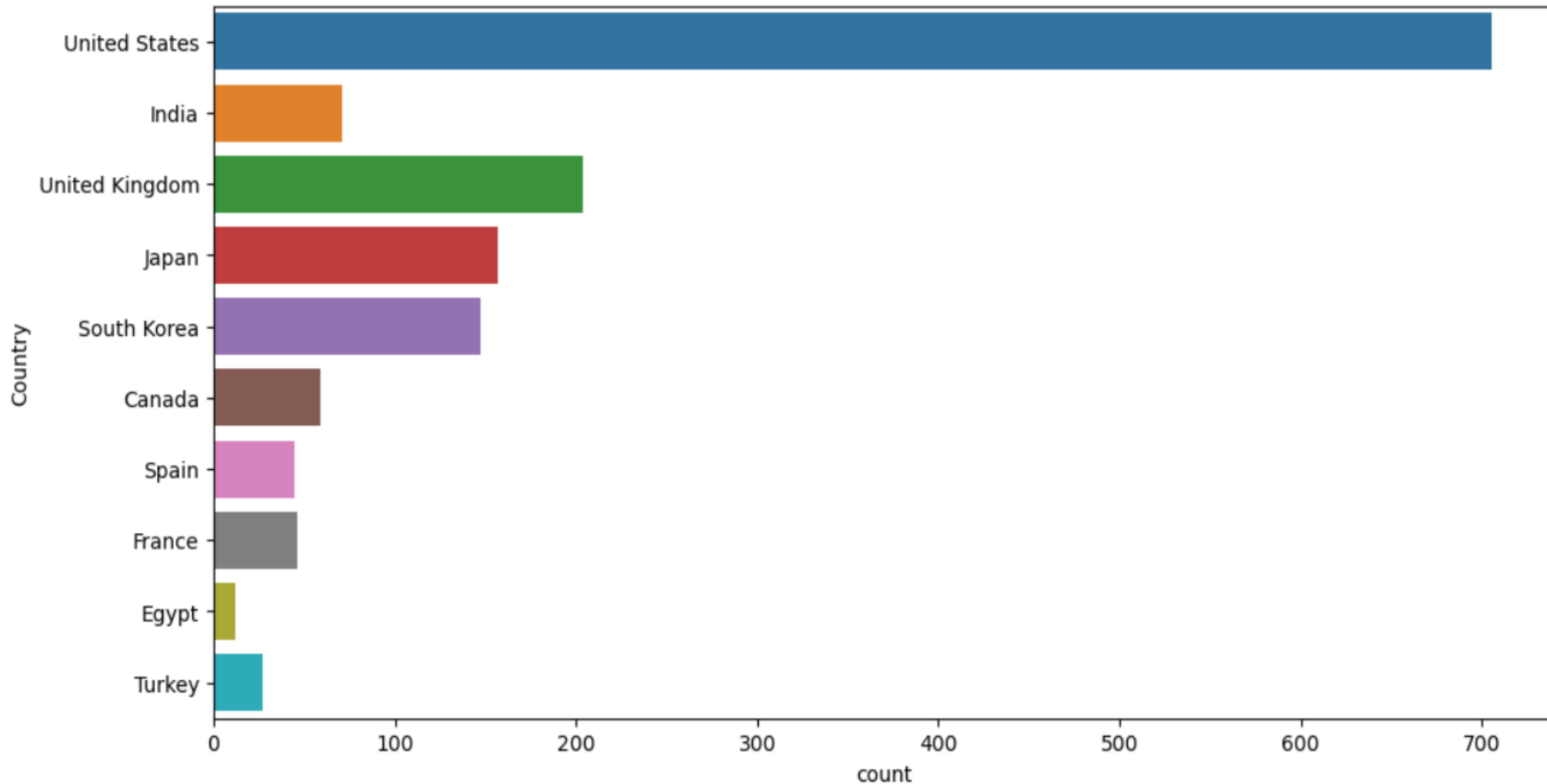
Checking Category Content based on Country

Text(0.5, 1.0, 'Top 10 Coutry producing TV Shows in Netflix')



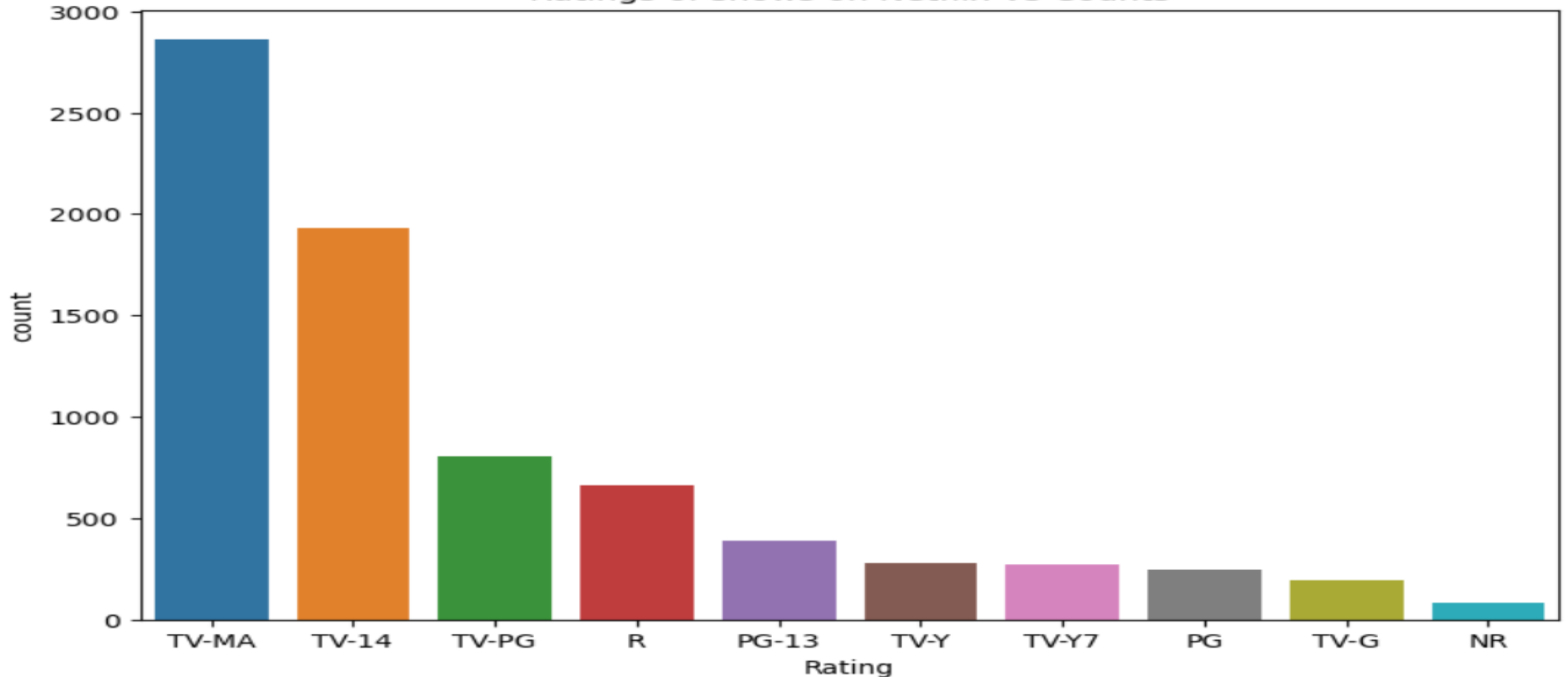
Top 10 Coutry producing TV Shows in Netflix

Top 10 Coutry producing TV Shows in Netflix

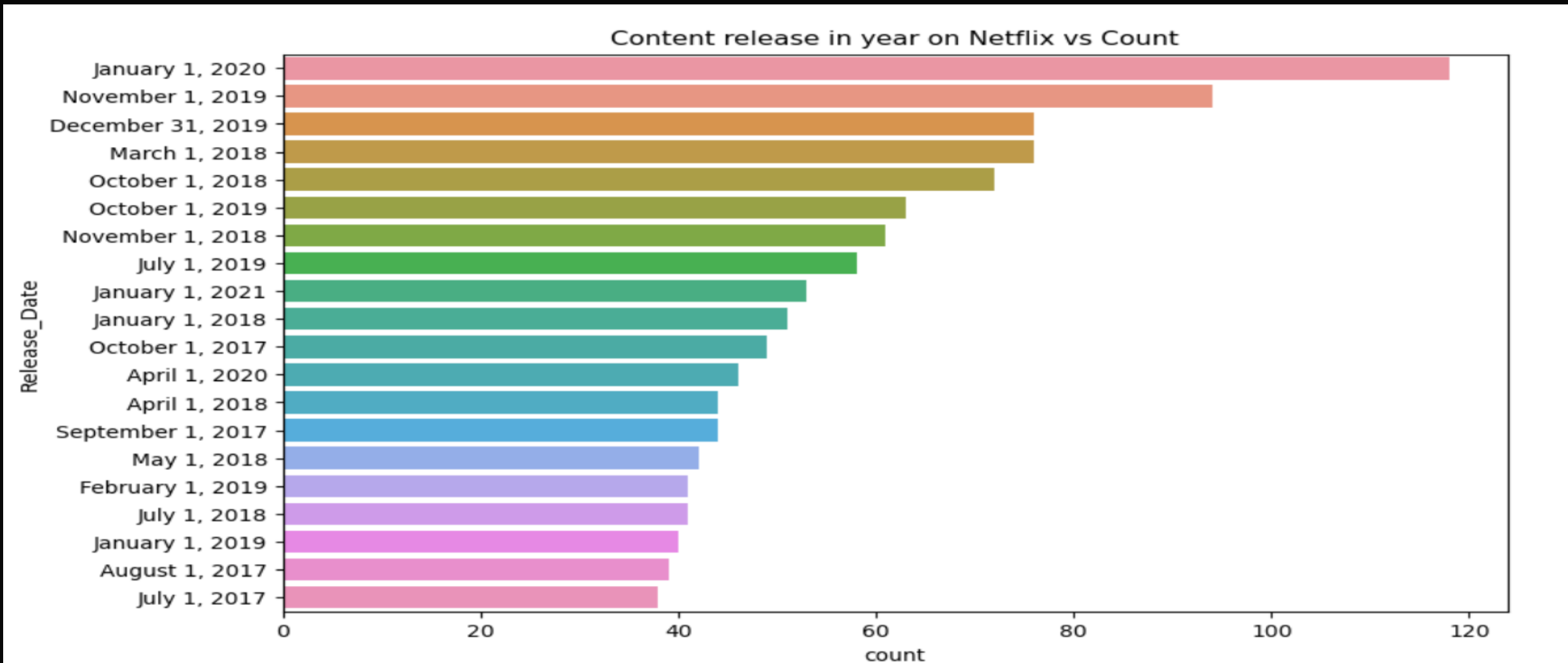


Checking major ratings given to Netflix Shows

Ratings of shows on Netflix Vs Counts



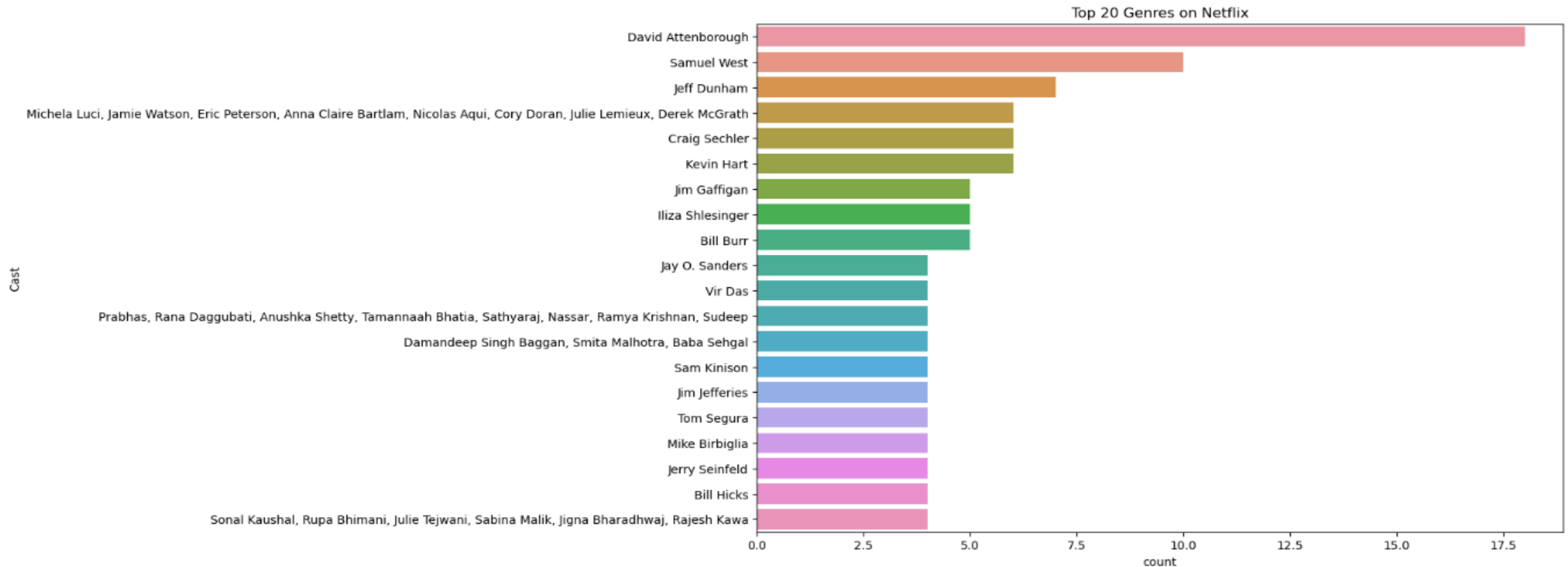
Most of the shows has TV-MA and TV-14 ratings



Finding Popular Cast Analysis

```
In [38]: plt.figure(figsize = (12,8))  
sns.countplot(y= 'Cast' ,order = data['Cast'].value_counts().index[0:20],data = data)  
plt.title('Top 20 Genres on Netflix')
```

Out[38]: Text(0.5, 1.0, 'Top 20 Genres on Netflix')

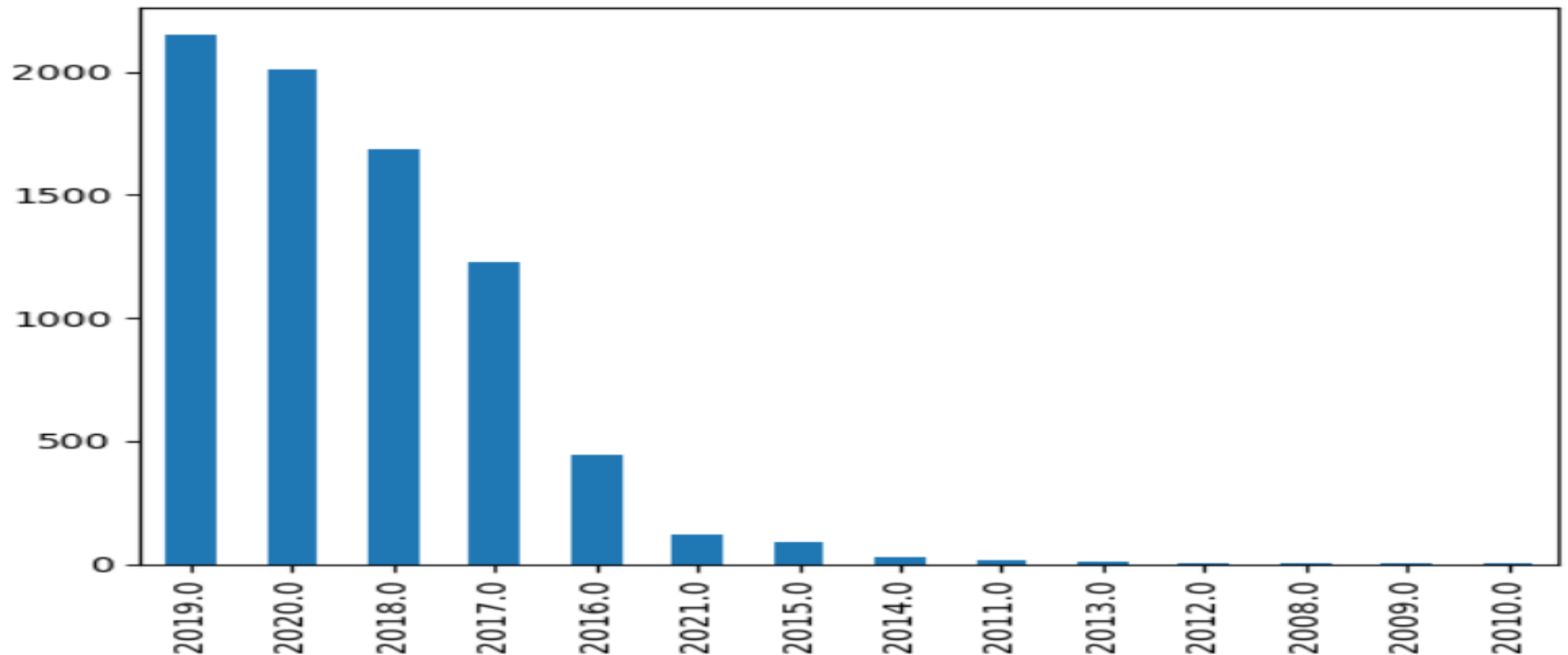


In which year the highest number of TV shows and Movies were Released? Showing with Bar Graph

Bar Graph

```
In [49]: data["Date_N"].dt.year.value_counts().plot(kind='bar')
```

```
Out[49]: <AxesSubplot:>
```



NETFLIX

THANK YOU