

```
In [1]: ▶ import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

In [2]: ▶ from wordcloud import WordCloud

In [3]: ▶ df=pd.read_csv('netflix.csv')
df
```

Out[3]:

| | 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, filmm... |
| 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, Alam 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 I... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8802 | s8803 | Movie | Zodiac | David Fincher | Mark Ruffalo, Jake Gyllenhaal, Robert Downey J... | United States | November 20, 2019 | 2007 | R | 158 min | Cult Movies, Dramas, Thrillers | A political cartoonist, a crime reporter and a... |
| 8803 | s8804 | TV Show | Zombie Dumb | NaN | NaN | NaN | July 1, 2019 | 2018 | TV-Y7 | 2 Seasons | Kids' TV, Korean TV Shows, TV Comedies | While living alone in a spooky town, a young g... |
| 8804 | s8805 | Movie | Zombieland | Ruben Fleischer | Jesse Eisenberg, Woody Harrelson, 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 | Mozez 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 × 12 columns

Basic Analysis

```
In [4]: ▶ df.shape

Out[4]: (8807, 12)
```

- Dataset is having 8807 rows and 12 columns

```
In [5]: df.info()
```

```
<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      8797 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
```

Handling null values

```
In [6]: df.isnull().sum()
```

```
Out[6]: 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 [7]: df['director'].fillna('Unknown director',inplace=True)
df
```

Out[7]:

| | 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, filmm... |
| 1 | s2 | TV Show | | Blood & Water | Unknown director | 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 | Unknown director | 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 | Unknown director | Mayur More, Jitendra Kumar, Ranjan Raj, Alam 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 I... |
| ... | ... | ... | | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8802 | s8803 | Movie | | Zodiac | David Fincher | Mark Ruffalo, Jake Gyllenhaal, Robert Downey J... | United States | November 20, 2019 | 2007 | R | 158 min | Cult Movies, Dramas, Thrillers | A political cartoonist, a crime reporter and a... |
| 8803 | s8804 | TV Show | | Zombie Dumb | Unknown director | NaN | NaN | July 1, 2019 | 2018 | TV-Y7 | 2 Seasons | Kids' TV, Korean TV Shows, TV Comedies | While living alone in a spooky town, a young g... |
| 8804 | s8805 | Movie | | Zombieland | Ruben Fleischer | Jesse Eisenberg, Woody Harrelson, 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 | Mozez 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 × 12 columns

```
In [8]: df['director'].fillna('Unknown director',inplace=True)
df['cast'].fillna('Unknown cast',inplace=True)
df['country'].fillna('Unknown country',inplace=True)
df['rating'].fillna('Unknown rating',inplace=True)
```

```
In [9]: continuous_columns = ['duration']
for column in continuous_columns:
    df[column].fillna(0, inplace=True)
```

```
In [10]: df['date_added'].dropna(inplace=True)
```

Dropped null values from date_added column.

```
In [11]: df.head(10)
```

Out[11]:

| | 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 | Unknown cast | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmm... |
| 1 | s2 | TV Show | Blood & Water | Unknown director | 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... | Unknown 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 lor... |
| 3 | s4 | TV Show | Jailbirds New Orleans | Unknown director | Unknown cast | Unknown 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 | Unknown director | Mayur More, Jitendra Kumar, Ranjan Raj, Alam 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 I... |
| 5 | s6 | TV Show | Midnight Mass | Mike Flanagan | Kate Siegel, Zach Gilford, Hamish Linklater, H... | Unknown country | September 24, 2021 | 2021 | TV-MA | 1 Season | TV Dramas, TV Horror, TV Mysteries | The arrival of a charismatic young priest brin... |
| 6 | s7 | Movie | My Little Pony: A New Generation | Robert Cullen, José Luis Ucha | Vanessa Hudgens, Kimiko Glenn, James Marsden, ... | Unknown country | September 24, 2021 | 2021 | PG | 91 min | Children & Family Movies | Equestria's divided. But a bright-eyed hero be... |
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | United States, Ghana, Burkina Faso, United Kin... | September 24, 2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... |
| 8 | s9 | TV Show | The Great British Baking Show | Andy Devonshire | Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho... | United Kingdom | September 24, 2021 | 2021 | TV-14 | 9 Seasons | British TV Shows, Reality TV | A talented batch of amateur bakers face off in... |
| 9 | s10 | Movie | The Starling | Theodore Melfi | Melissa McCarthy, Chris O'Dowd, Kevin Kline, T... | United States | September 24, 2021 | 2021 | PG-13 | 104 min | Comedies, Dramas | A woman adjusting to life after a loss contend... |

What does ‘good’ look like?

The count of each categorical variable both using graphical and non-graphical analysis.

Non graphical analysis

```
In [41]: df['listed_in'].
```

Out[41]: 0 Documentaries
1 International TV Shows, TV Dramas, TV Mysteries
2 Crime TV Shows, International TV Shows, TV Act...
3 Docuseries, Reality TV
4 International TV Shows, Romantic TV Shows, TV ...
...
8802 Cult Movies, Dramas, Thrillers
8803 Kids' TV, Korean TV Shows, TV Comedies
8804 Comedies, Horror Movies
8805 Children & Family Movies, Comedies
8806 Dramas, International Movies, Music & Musicals
Name: listed_in, Length: 10835, dtype: object

In []: ▶

In []: ▶

In [12]: ▶ `df['type'].value_counts()`

Out[12]: type
Movie 6131
TV Show 2676
Name: count, dtype: int64

- Clearly Movies content is almost three times more compared to TV show content.(i.e. 6131 movies & 2676 TV Show)

In [13]: ▶ `actors= df['cast'].str.split(',')
all_actor=[actor.lstrip() for sublist in actors for actor in sublist]
df_actor=pd.DataFrame(all_actor)
all_the_actors=df_actor.value_counts()
all_the_actors_df= pd.DataFrame(all_the_actors)
all_the_actors_df.reset_index(inplace=True)
all_the_actors_df.rename(columns={0: 'cast'},inplace=True)
all_the_actors_df`

Out[13]:

| | cast | count |
|-------|------------------|-------|
| 0 | Unknown cast | 825 |
| 1 | Anupam Kher | 43 |
| 2 | Shah Rukh Khan | 35 |
| 3 | Julie Teiwani | 33 |
| 4 | Naseeruddin Shah | 32 |
| ... | ... | ... |
| 36435 | Jamie Lee | 1 |
| 36436 | Jamie Kenna | 1 |
| 36437 | Jamie Kaler | 1 |
| 36438 | Jamie Johnston | 1 |
| 36439 | Şopé Dirisù | 1 |

36440 rows × 2 columns

In [14]: ▶ `directors= df['director'].str.split(',')
all_directors= [any_director.lstrip() for sublist in directors for any_director in sublist]
df_directors=pd.DataFrame(all_directors)
df_all_directors=df_directors.value_counts()
df_all_the_directors=pd.DataFrame(df_all_directors)
df_all_the_directors.reset_index(inplace=True)
df_all_the_directors.rename(columns={0: 'director'},inplace=True)
df_all_the_directors`

Out[14]:

| | director | count |
|------|-------------------------|-------|
| 0 | Unknown director | 2634 |
| 1 | Rajiv Chilaka | 22 |
| 2 | Jan Suter | 21 |
| 3 | Raúl Campos | 19 |
| 4 | Marcus Raboy | 16 |
| ... | ... | ... |
| 4989 | Juan Antin | 1 |
| 4990 | Juan Antonio de la Riva | 1 |
| 4991 | Juan Camilo Pinzon | 1 |
| 4992 | Juan Carlos Medina | 1 |
| 4993 | Ayush Raina | 1 |

In [15]: ▶ `countries=df['country'].str.split(',')
all_countries=[any_country.lstrip() for sublist in countries for any_country in sublist]
d_countries=pd.DataFrame(all_countries).value_counts()
df_countries=d_countries.to_frame().reset_index()
df_countries.rename(columns={0: 'country'},inplace=True)`

```
In [17]: df_countries
```

Out[17]:

| | country | count |
|-----|-----------------|-------|
| 0 | United States | 3690 |
| 1 | India | 1046 |
| 2 | Unknown country | 831 |
| 3 | United Kingdom | 806 |
| 4 | Canada | 445 |
| ... | ... | ... |
| 119 | Jamaica | 1 |
| 120 | Slovakia | 1 |
| 121 | Somalia | 1 |
| 122 | Ethiopia | 1 |
| 123 | Lithuania | 1 |

124 rows × 2 columns

- United States has dominated the content release followed by India and United Kingdom.

```
In [18]: df['release_year'].value_counts()
```

Out[18]:

| release_year | count |
|--------------|-------|
| 2018 | 1147 |
| 2017 | 1032 |
| 2019 | 1030 |
| 2020 | 953 |
| 2016 | 902 |
| ... | ... |
| 1959 | 1 |
| 1925 | 1 |
| 1961 | 1 |
| 1947 | 1 |
| 1966 | 1 |

Name: count, Length: 74, dtype: int64

```
In [19]: df['rating'].unique()
```

Out[19]:

```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',  
      'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR',  
      'Unknown rating', 'TV-Y7-FV', 'UR'], dtype=object)
```

```
In [20]: rating_data=df['rating'].value_counts()  
rating_df=rating_data.to_frame().reset_index()  
rating_df
```

| | | |
|----|----------------|-----|
| 5 | TV-Y7 | 334 |
| 6 | TV-Y | 307 |
| 7 | PG | 287 |
| 8 | TV-G | 220 |
| 9 | NR | 80 |
| 10 | G | 41 |
| 11 | TV-Y7-FV | 6 |
| 12 | Unknown rating | 4 |
| 13 | NC-17 | 3 |
| 14 | UR | 3 |
| 15 | 74 min | 1 |
| 16 | 84 min | 1 |
| 17 | 66 min | 1 |

```
In [21]: df['country'] = df['country'].str.split(', ')
df = df.explode('country')
country_wise_rating=df.groupby('country')['rating'].value_counts()
df_country_wise_rating=country_wise_rating.to_frame()
df_country_wise_rating.reset_index(inplace=True)
df_country_wise_rating
```

Out[21]:

| | country | rating | count |
|-----|--------------|--------|-------|
| 0 | | TV-14 | 1 |
| 1 | | TV-MA | 1 |
| 2 | Afghanistan | TV-MA | 1 |
| 3 | Albania | TV-MA | 1 |
| 4 | Algeria | TV-MA | 2 |
| ... | ... | ... | ... |
| 528 | West Germany | TV-14 | 1 |
| 529 | West Germany | G | 1 |
| 530 | Zimbabwe | TV-G | 1 |
| 531 | Zimbabwe | NR | 1 |
| 532 | Zimbabwe | TV-MA | 1 |

533 rows × 3 columns

```
In [22]: listedin=df['listed_in'].str.split(',')
all_category=[any_category.lstrip() for sublist in listedin for any_category in sublist ]
categories_df= pd.DataFrame(all_category)
categories_df.reset_index(inplace=True)
categories_df.rename(columns={0:'genre'},inplace=True)
categories_df
```

| | | |
|-------|-------|--------------------------|
| 0 | 0 | Documentaries |
| 1 | 1 | International TV Shows |
| 2 | 2 | TV Dramas |
| 3 | 3 | TV Mysteries |
| 4 | 4 | Crime TV Shows |
| ... | ... | ... |
| 23749 | 23749 | Children & Family Movies |
| 23750 | 23750 | Comedies |
| 23751 | 23751 | Dramas |
| 23752 | 23752 | International Movies |
| 23753 | 23753 | Music & Musicals |

23754 rows × 2 columns

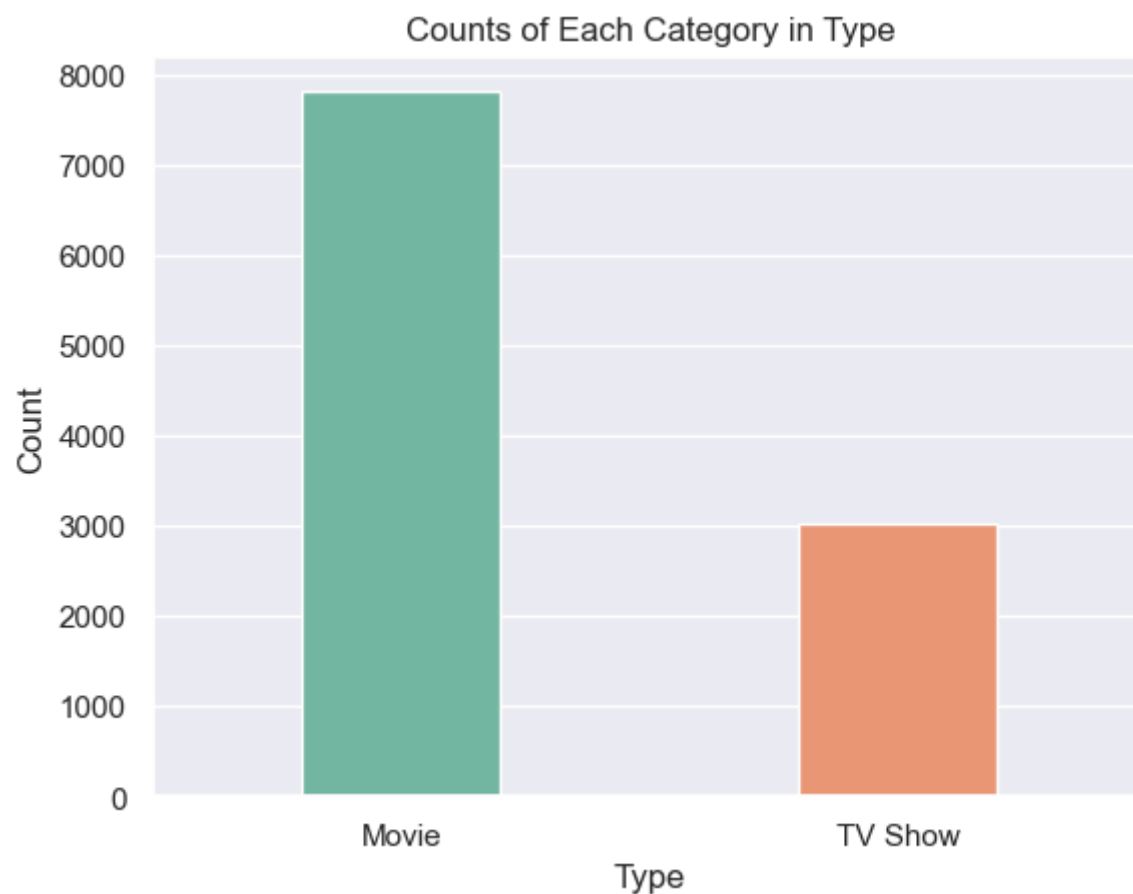
```
In [23]: movies = df.loc[df['type'] == 'Movie']
tv_shows = df.loc[df['type'] == 'TV Show']
```

```
In [24]: movies.duration.value_counts()
```

```
Out[24]: duration
93 min    202
90 min    199
94 min    187
96 min    187
97 min    183
...
203 min     1
253 min     1
230 min     1
237 min     1
191 min     1
Name: count, Length: 206, dtype: int64
```

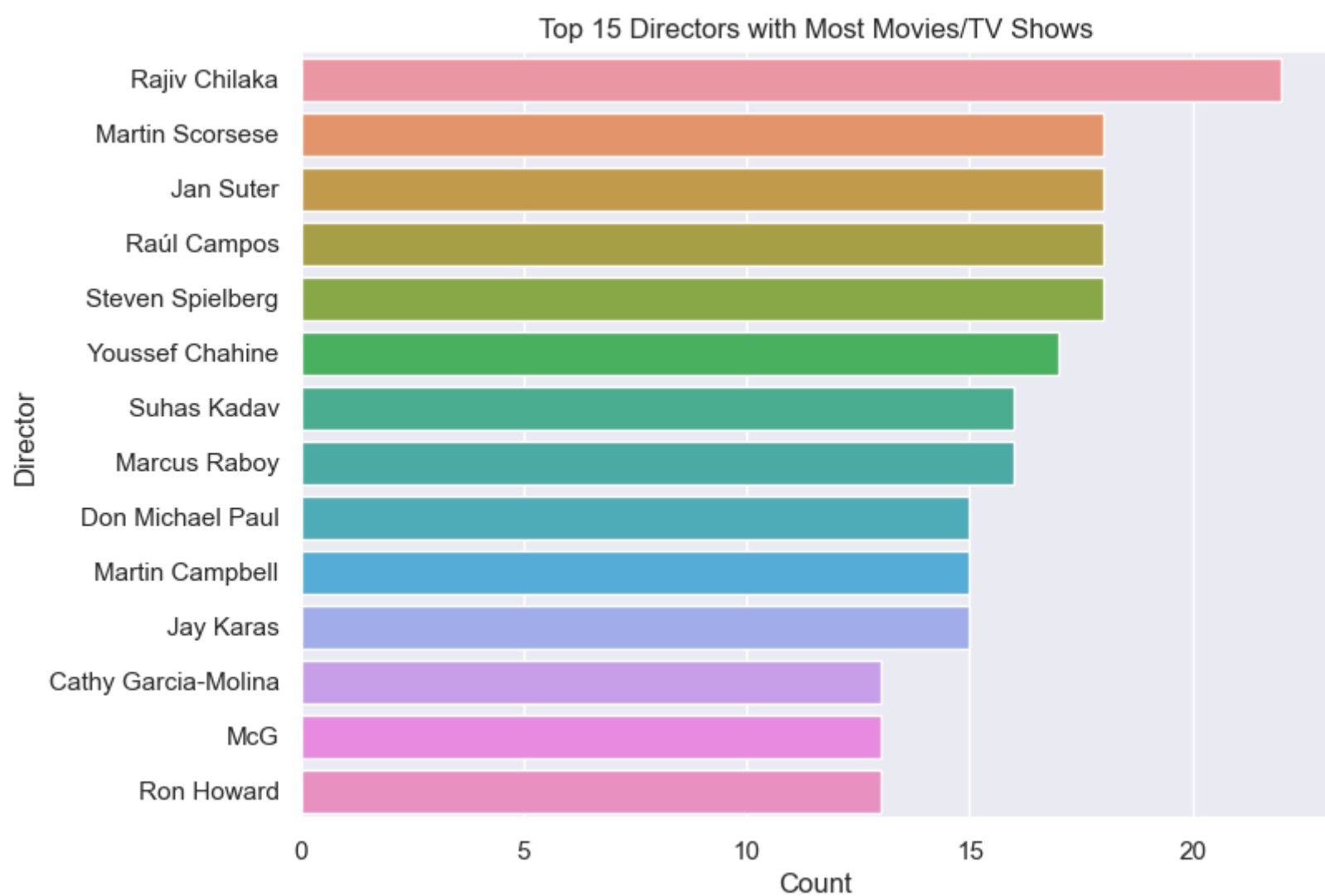
Graphical analysis

```
In [80]: ▶ sns.countplot(x='type', data=df,width=0.4, palette="Set2")
sns.set(style="darkgrid")
plt.title('Counts of Each Category in Type')
plt.xlabel('Type')
plt.ylabel('Count')
plt.savefig('Counts of Each Category in Type.jpg')
plt.show()
```



```
In [84]: ▶ all_directors= [any_director for sublist in directors for any_director in sublist]
all_directors_df=pd.DataFrame(all_directors)
all_directors_df.reset_index()
all_directors_df.rename(columns={0:'directors'},inplace=True)
all_directors_df
plt.figure(figsize=(8, 6))
sns.countplot(y='directors', data=all_directors_df, order=all_directors_df['directors'].value_counts().index[1:15])

plt.title('Top 15 Directors with Most Movies/TV Shows')
plt.xlabel('Count')
plt.ylabel('Director')
plt.savefig('top 15 directors')
plt.show()
```



Year Wise content history

```
In [27]: df.drop(df.loc[df['date_added'].isna()].index , axis = 0 , inplace = True)
```

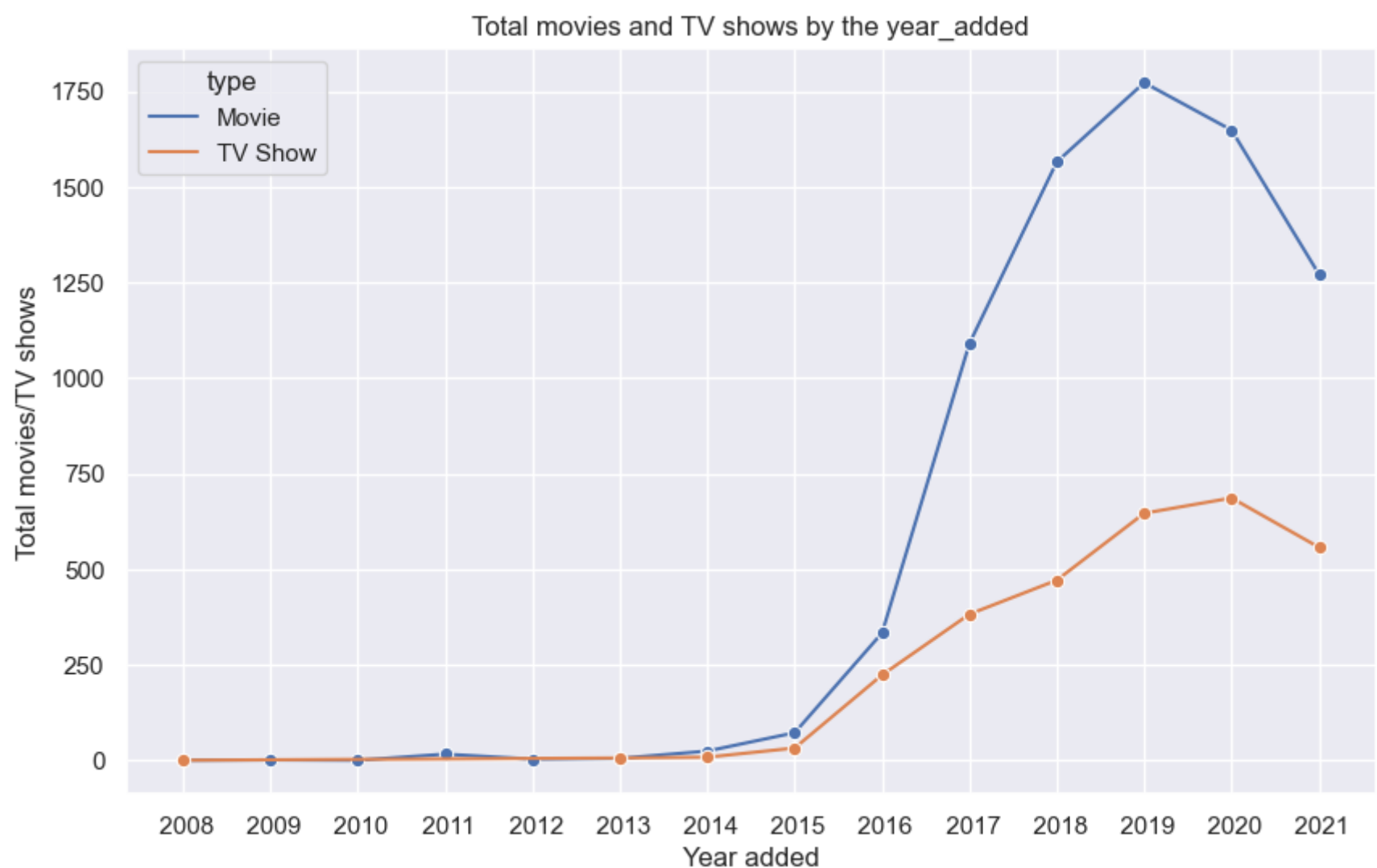
```
In [28]: df['date_added'].value_counts()
```

```
Out[28]: date_added
January 1, 2020      164
November 1, 2019    108
July 1, 2021         92
December 31, 2019   83
March 1, 2018        81
...
December 12, 2016     1
December 30, 2016     1
January 3, 2017        1
January 4, 2017        1
January 11, 2020       1
Name: count, Length: 1767, dtype: int64
```

```
In [29]: df['year'] = df['date_added'].apply(lambda x:x.split(',')[0])
df['month'] = df['date_added'].apply(lambda x:x.split(',')[1])
```

```
In [30]: d = df.groupby(['year' , 'type' ])[ 'show_id' ].count().reset_index()
d.rename({'show_id' : 'total movies/TV shows'}, axis = 1 , inplace = True)
```

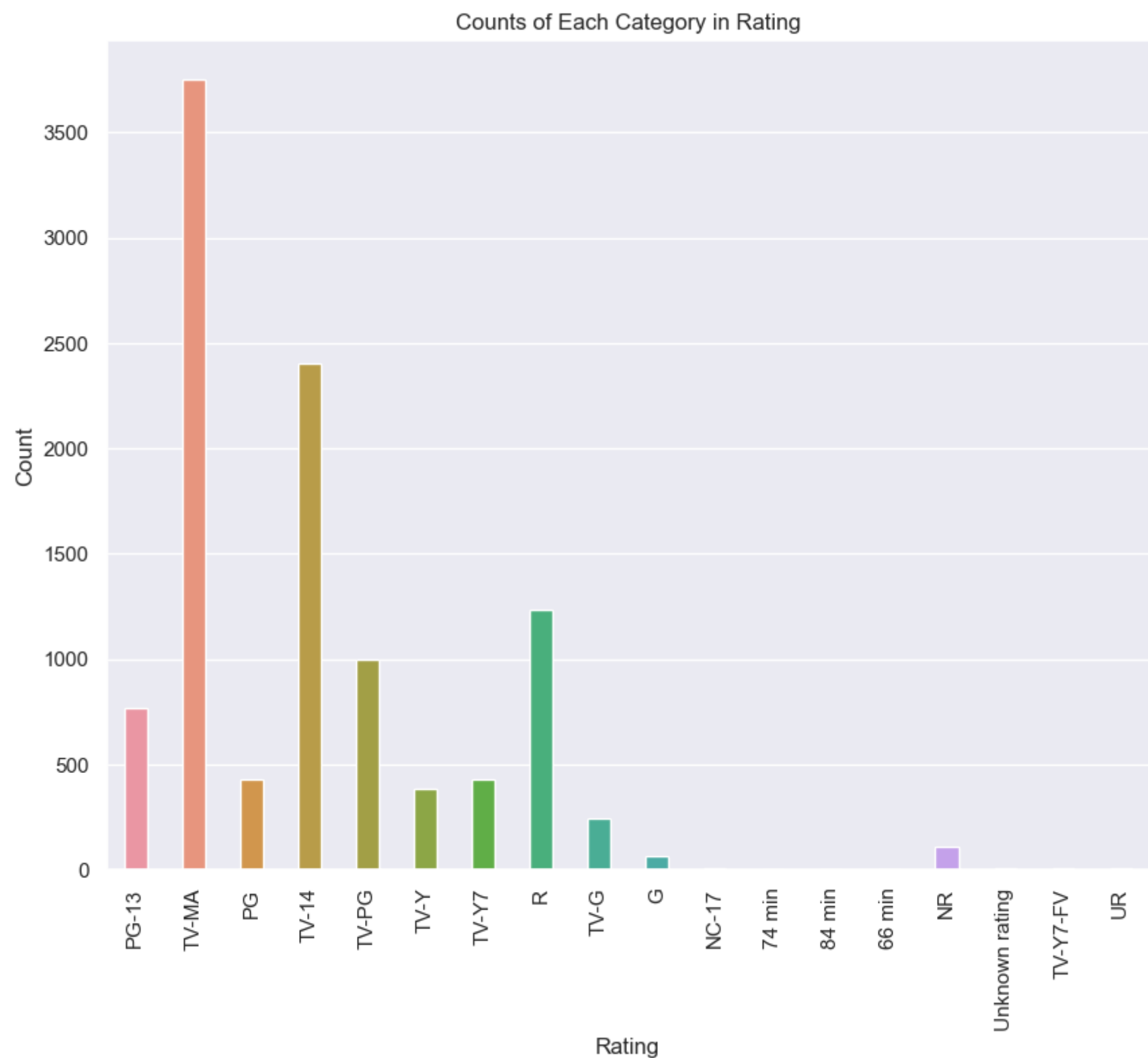
```
In [85]: plt.figure(figsize=(10,6))
sns.lineplot(data=d,x='year',y='total movies/TV shows', hue='type', marker='o',ms=6)
plt.xlabel('Year added')
plt.ylabel('Total movies/TV shows')
plt.title('Total movies and TV shows by the year_added' , fontsize = 12)
plt.savefig('total movies and TV show by year')
plt.show()
```



Observation:

- The content added on the Netflix surged drastically after 2015.
- 2019 marks the highest number of movies and TV shows added on the Netflix.
- Year 2020 and 2021 has seen the drop in content added on Netflix, possibly because of Pandemic. But still , TV shows content have not dropped as drastic as movies. In recent years TV shows are focussed more than Movies.

```
In [86]: ▶ plt.figure(figsize=(10,8))
sns.countplot(x='rating', data=df,width=0.4)
plt.title('Counts of Each Category in Rating')
plt.xticks(rotation=90)
plt.xlabel('Rating')
plt.ylabel('Count')
plt.savefig('Counts of each category in rating')
plt.show()
```



- Most content available on Netflix is for matured audience followed by

```
In [35]: ▶ categories_df
```

Out[35]:

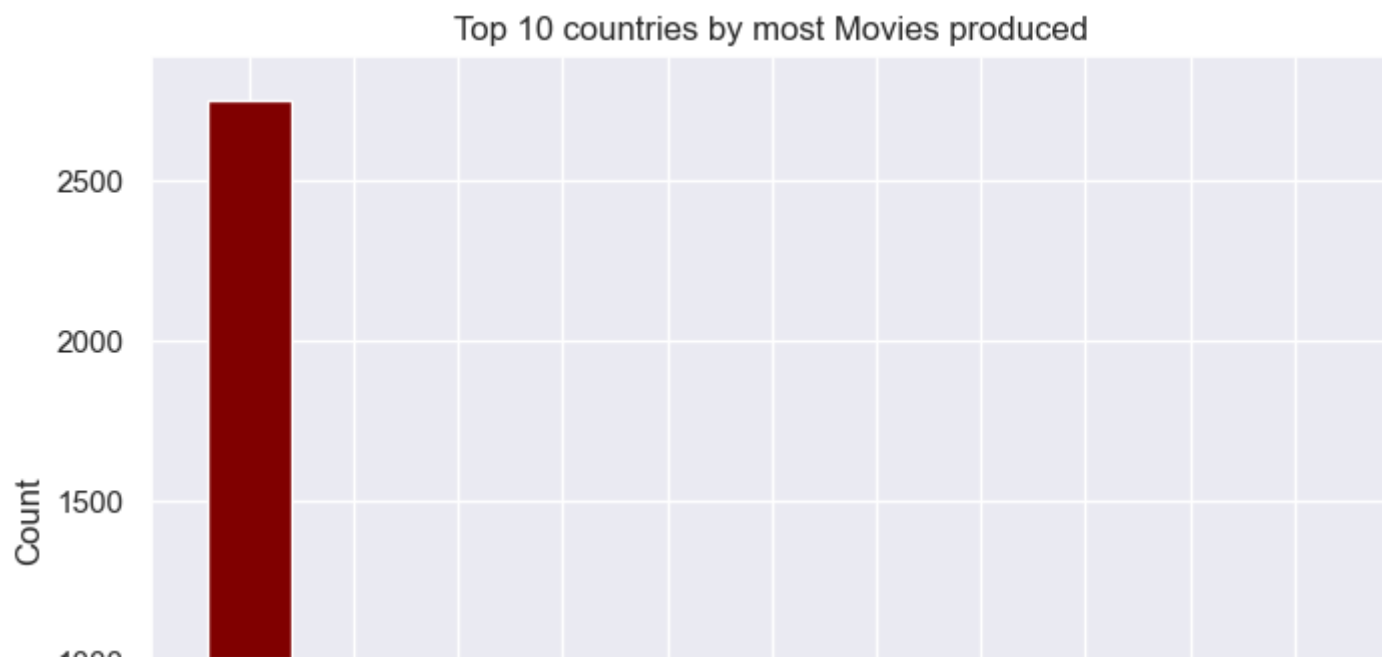
| | index | genre |
|-------|-------|--------------------------|
| 0 | 0 | Documentaries |
| 1 | 1 | International TV Shows |
| 2 | 2 | TV Dramas |
| 3 | 3 | TV Mysteries |
| 4 | 4 | Crime TV Shows |
| ... | ... | ... |
| 23749 | 23749 | Children & Family Movies |
| 23750 | 23750 | Comedies |
| 23751 | 23751 | Dramas |
| 23752 | 23752 | International Movies |
| 23753 | 23753 | Music & Musicals |

23754 rows × 2 columns

Comparison of tv shows vs. movies

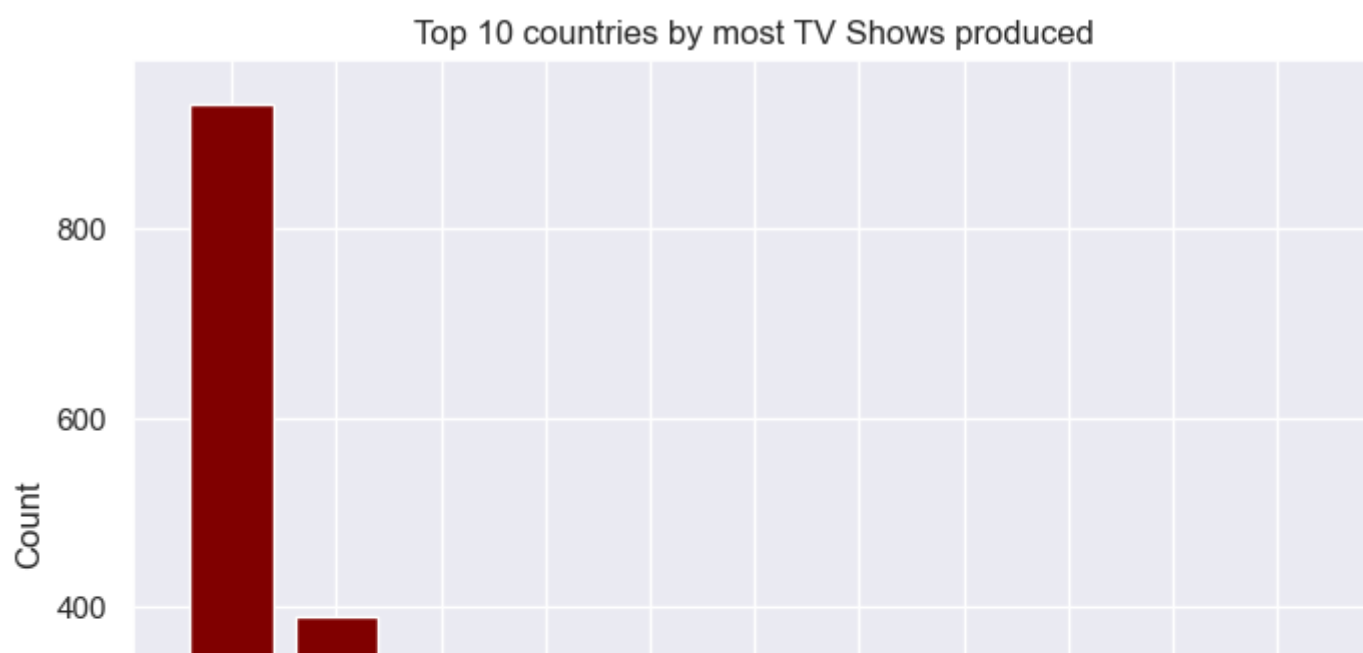
Top 10 countries by most movies produced

```
In [87]: ▶ movies_data=df[df['type']=='Movie']
top_country_movies= movies_data.groupby('country')['type'].value_counts().sort_values(ascending=False).reset_index()
y=top_country_movies['count']
x=top_country_movies['country']
plt.figure(figsize=(8,6))
plt.bar(x,y,data=movies_data,color='maroon')
plt.title('Top 10 countries by most Movies produced')
plt.xticks(rotation=90)
plt.xlabel('Countries')
plt.ylabel('Count')
plt.savefig('top 10 countries by most movies produced')
plt.show()
```



Top 10 countries by most TV Shows produced

```
In [88]: ▶ tv_shows_data=df[df['type']=='TV Show']
top_country_TV_show=tv_shows_data.groupby('country')['type'].value_counts().sort_values(ascending=False).head(11)
y=top_country_TV_show['count']
x=top_country_TV_show['country']
plt.figure(figsize=(8,6))
plt.bar(x,y,data=tv_shows_data,color='maroon')
plt.title('Top 10 countries by most TV Shows produced')
plt.xticks(rotation=90)
plt.xlabel('Countries')
plt.ylabel('Count')
plt.savefig('top 10 countries by tv shows')
plt.show()
```



What is the best time to launch a TV show?

```
In [45]: df_shows= df[df['type']=='TV Show']
df_shows.head()
```

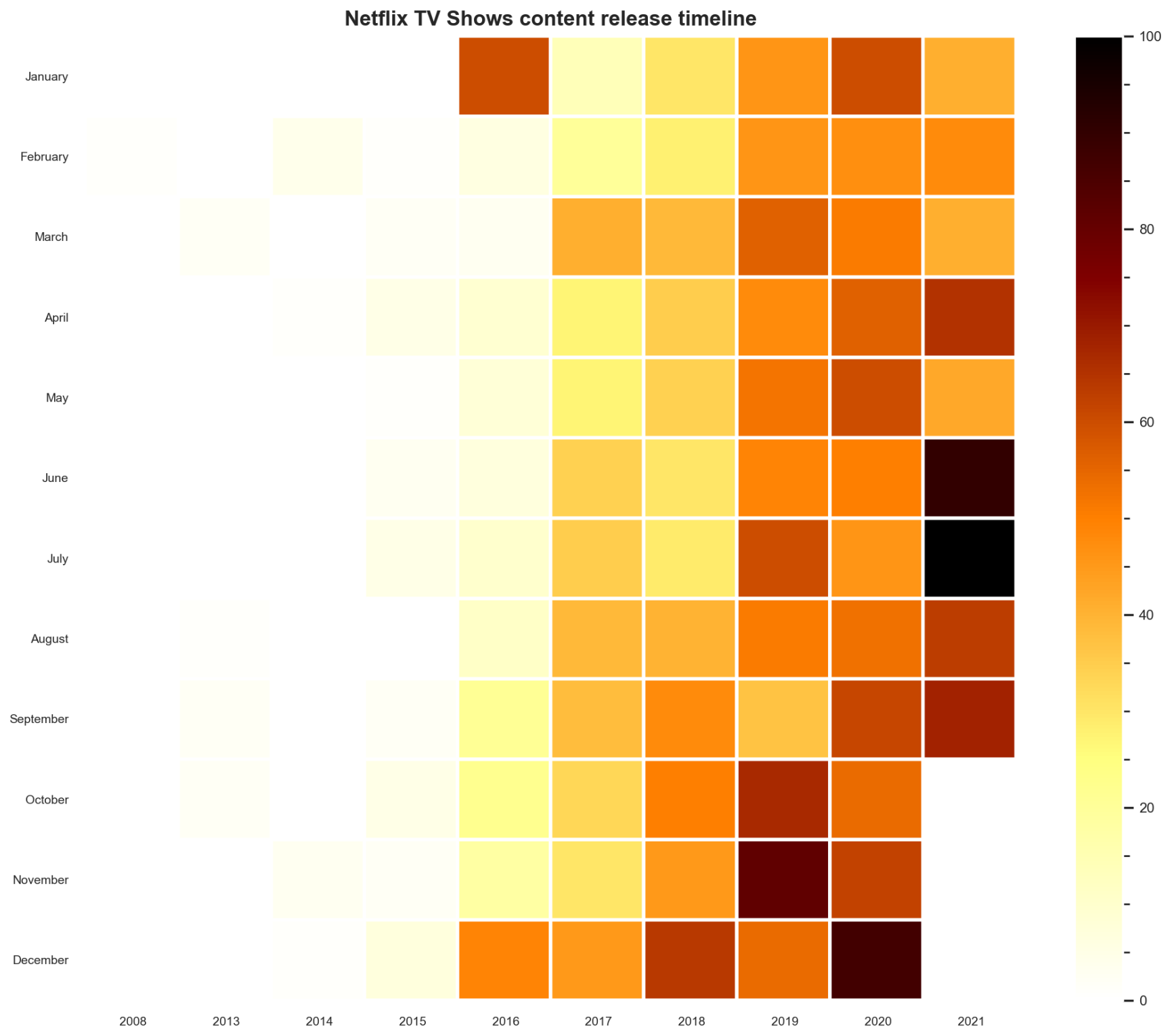
Out[45]:

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | year | |
|---|---------|---------|-----------------------|------------------|---|-----------------|--------------------|--------------|--------|-----------|---|---|------|------|
| 1 | s2 | TV Show | Blood & Water | Unknown director | 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... | 2021 | Sept |
| 2 | s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi... | Unknown 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 lor... | 2021 | Sept |
| 3 | s4 | TV Show | Jailbirds New Orleans | Unknown director | Unknown cast | Unknown country | September 24, 2021 | 2021 | TV-MA | 1 Season | Docuseries, Reality TV | Feuds, flirtations and toilet talk go down amo... | 2021 | Sept |
| 4 | s5 | TV Show | Kota Factory | Unknown director | Mayur More, Jitendra Kumar, Ranjan Raj, Alam 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... | 2021 | Sept |
| 5 | s6 | TV Show | Midnight Mass | Mike Flanagan | Kate Siegel, Zach Gilford, Hamish Linklater, H... | Unknown country | September 24, 2021 | 2021 | TV-MA | 1 Season | TV Dramas, TV Horror, TV Mysteries | The arrival of a charismatic young priest brin... | 2021 | Sept |

```
In [46]: df_shows_date = df_shows[['date_added']].dropna()
```

```
In [89]: ▶ df_shows_date['year'] = df_shows_date['date_added'].apply(lambda x:x.split(',')[1])
df_shows_date['month'] = df_shows_date['date_added'].apply(lambda x:x.lstrip().split(' ')[0])
month_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December']
dff=df_shows_date.groupby('year')['month'].value_counts().unstack().fillna(0)[month_order].T
plt.figure(figsize=(12,10),dpi=200)
plt.pcolor(dff, cmap='afmhot_r', edgecolors='white', linewidths=2)
plt.xticks(np.arange(0.5, len(dff.columns), 1), dff.columns, fontsize=7)
plt.yticks(np.arange(0.5, len(dff.index), 1), dff.index, fontsize=7)
plt.title('Netflix TV Shows content release timeline',fontweight='bold')
cbar = plt.colorbar()

cbar.ax.tick_params(labelsize=8)
cbar.ax.minorticks_on()
plt.savefig('best time to launch tvshow')
plt.show()
```



- If the latest year as 2020 is considered, it is evident that in month of February and June much less content was released as compared to other months. So this period seems was suitable for content release.

What is the best time to launch a Movie?

```
In [48]: df_movies= df[df['type']=='Movie']
df_movies.head()
```

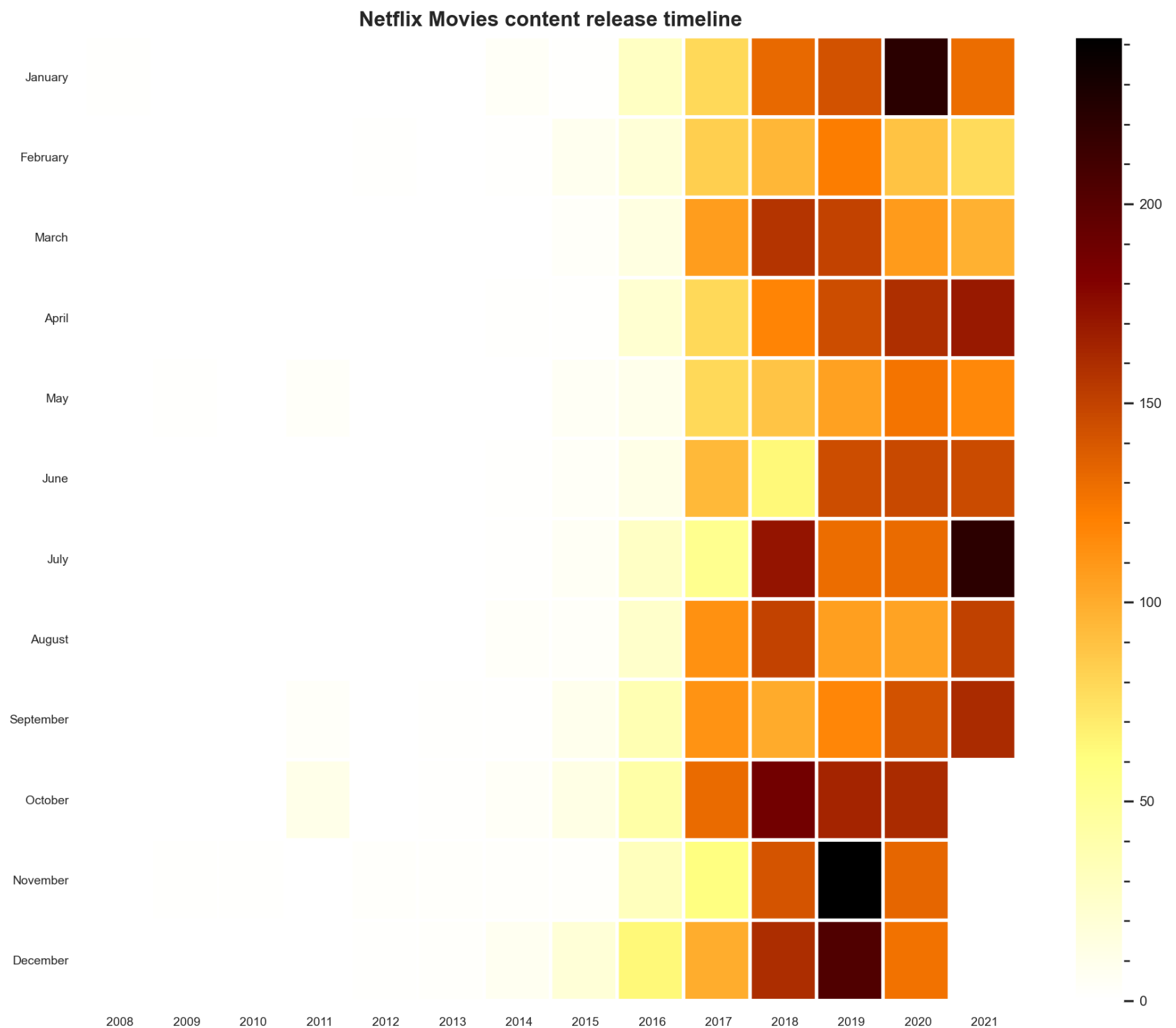
Out[48]:

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | year | |
|---|---------|-------|----------------------------------|-------------------------------|---|-----------------|--------------------|--------------|--------|----------|--|---|------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | Unknown cast | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmm... | 2021 | 5 |
| 6 | s7 | Movie | My Little Pony: A New Generation | Robert Cullen, José Luis Ucha | Vanessa Hudgens, Kimiko Glenn, James Marsden, ... | Unknown country | September 24, 2021 | 2021 | PG | 91 min | Children & Family Movies | Equestria's divided. But a bright-eyed hero be... | 2021 | 5 |
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | United States | September 24, 2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... | 2021 | 5 |
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | Ghana | September 24, 2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... | 2021 | 5 |
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | Burkina Faso | September 24, 2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... | 2021 | 5 |

```
In [49]: df_movies_date = df_movies[['date_added']].dropna()
```

```
In [95]: ▶ df_movies_date['year'] = df_movies_date['date_added'].apply(lambda x:x.split(',')[1])
df_movies_date['month'] = df_movies_date['date_added'].apply(lambda x:x.lstrip().split(' ')[0])
month_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December']
dff= df_movies_date.groupby('year')['month'].value_counts().unstack().fillna(0)[month_order].T
plt.figure(figsize=(12,10),dpi=200)
plt.pcolor(dff, cmap='afmhot_r', edgecolors='white', linewidths=2)
plt.xticks(np.arange(0.5, len(dff.columns), 1), dff.columns, fontsize=7)
plt.yticks(np.arange(0.5, len(dff.index), 1), dff.index, fontsize=7)
plt.title('Netflix Movies content release timeline',fontweight='bold')
cbar = plt.colorbar()

cbar.ax.tick_params(labelsize=8)
cbar.ax.minorticks_on()
plt.savefig('best time to add movie')
plt.show()
```



- If the latest year as 2020 is considered, it is evident that in month of February and August, much less content was released as compared to other months. So this period was much suitable for content release

Analysis of actors/directors of different types of shows/movies.

Identify the top 10 actors who have appeared in most movies or TV shows.

```
In [51]: ▶ actors= df['cast'].str.split(',')

```

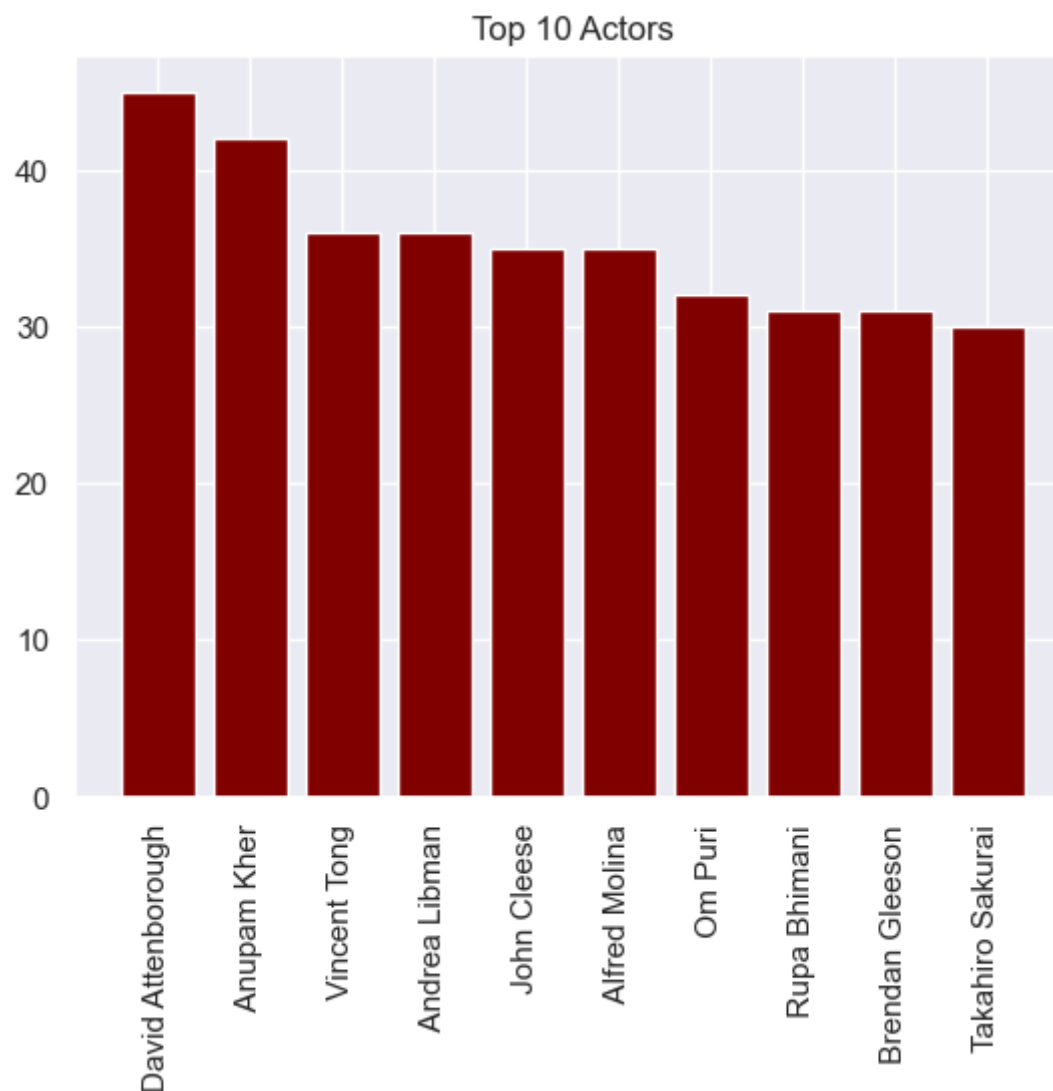
```
In [52]: ▶ all_actor=[actor for sublist in actors for actor in sublist]
all_actor
len(all_actor)
df_actor=pd.DataFrame(all_actor)
top_10_actors=df_actor.value_counts().sort_values(ascending=False)[1:11].to_frame().reset_index()
top_10_actors.rename(columns={0:'actors'},inplace=True)
top_10_actors
```

Out[52]:

| | actors | count |
|---|--------------------|-------|
| 0 | David Attenborough | 45 |
| 1 | Anupam Kher | 42 |
| 2 | Vincent Tong | 36 |
| 3 | Andrea Libman | 36 |
| 4 | John Cleese | 35 |
| 5 | Alfred Molina | 35 |
| 6 | Om Puri | 32 |
| 7 | Rupa Bhimani | 31 |
| 8 | Brendan Gleeson | 31 |
| 9 | Takahiro Sakurai | 30 |

```
In [53]: ▶ plt.bar(data=top_10_actors,x='actors',height='count',color='maroon')
plt.xticks(rotation=90)
plt.title('Top 10 Actors')

plt.show()
```



Identify the top 10 directors who have appeared in most movies or TV shows.

```
In [54]: ▶ directors= df['director'].str.split(',')
directors
```

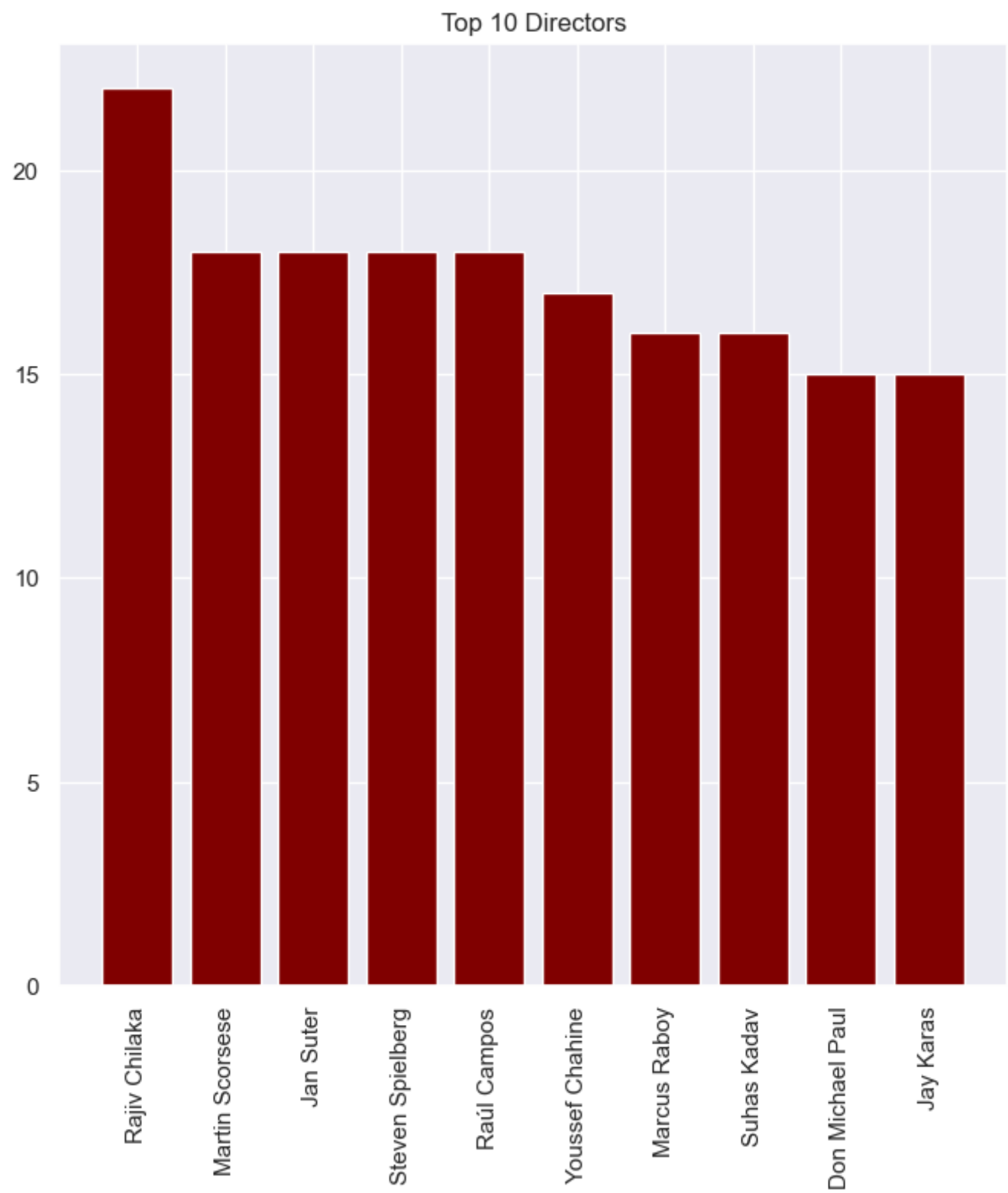
```
In [55]: ▶ all_directors= [any_director for sublist in directors for any_director in sublist]
all_directors
len(all_directors)
```

Out[55]: 11902

```
In [56]: ▶ df_directors=pd.DataFrame(all_directors)
top_10_directors=df_directors.value_counts().sort_values(ascending=False)[1:11].to_frame().reset_index()
top_10_directors.rename(columns={0:'directors'},inplace=True)
```



```
In [57]: plt.figure(figsize=(8,8))
plt.bar(data=top_10_directors,x='directors',height='count',color='maroon')
plt.title('Top 10 Directors')
plt.xticks(rotation=90)
plt.show()
```



Which genre movies are more popular or produced more?

```
In [58]: movies_df=df[df['type']=='Movie']
movies_df.head()
```

| | | | | | | | | | | | | | |
|---|----|-------|----------------------------------|-------------------------------|---|-----------------|--------------------|------|-------|---------|--|---|-----|
| 6 | s7 | Movie | My Little Pony: A New Generation | Robert Cullen, José Luis Ucha | Vanessa Hudgens, Kimiko Glenn, James Marsden, ... | Unknown country | September 24, 2021 | 2021 | PG | 91 min | Children & Family Movies | Equestria's divided. But a bright-eyed hero be... | 202 |
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | United States | September 24, 2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... | 202 |
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | Ghana | September 24, 2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... | 202 |
| | | | | | Kofi | | | | | | Dramas. | On a photo | |

```
In [59]: category=movies_df['listed_in'].str.split(',')
all_category=[anycategory.lstrip() for sublist in category for anycategory in sublist]
all_category
```

```
Documentaries ,
'Sports Movies',
'Dramas',
'International Movies',
'Thrillers',
'Action & Adventure',
'Comedies',
'Dramas',
'Children & Family Movies',
'Dramas',
'Dramas',
'Comedies',
'Romantic Movies',
'Dramas',
'International Movies',
'Thrillers',
'Dramas',
'Romantic Movies',
'Dramas',
'Romantic Movies',
,
```

```
In [60]: categories= pd.DataFrame(all_category)
categories.rename(columns={0:'categories'},inplace=True)
categories_df=categories.value_counts()
all_categories_df=pd.DataFrame(categories_df)
all_categories_df.reset_index(inplace=True)
all_categories_df
#categories.value_counts()
```

Out[60]:

| | categories | count |
|----|--------------------------|-------|
| 0 | International Movies | 3509 |
| 1 | Dramas | 3199 |
| 2 | Comedies | 1981 |
| 3 | Action & Adventure | 1182 |
| 4 | Documentaries | 1116 |
| 5 | Independent Movies | 1040 |
| 6 | Children & Family Movies | 845 |
| 7 | Thrillers | 806 |
| 8 | Romantic Movies | 722 |
| 9 | Horror Movies | 458 |
| 10 | Music & Musicals | 429 |
| 11 | Sci-Fi & Fantasy | 380 |
| 12 | Stand-Up Comedy | 344 |
| 13 | Sports Movies | 265 |
| 14 | Classic Movies | 151 |
| 15 | LGBTQ Movies | 116 |
| 16 | Cult Movies | 93 |
| 17 | Faith & Spirituality | 89 |
| 18 | Anime Features | 78 |
| 19 | Movies | 70 |

```
In [61]: all_genres = ' '.join(all_category)
```

```
In [62]: all_genres
cs international movies romantic movies comedies international movies romantic movies action & adventure inte
rnational Movies Stand-Up Comedy Comedies Dramas International Movies Documentaries International Movies Docu
mentaries International Movies Music & Musicals Dramas International Movies Thrillers Dramas Faith & Spiritua
lity Romantic Movies Dramas Independent Movies LGBTQ Movies Action & Adventure Comedies Independent Movies LG
BTQ Movies Action & Adventure Cult Movies Sci-Fi & Fantasy Documentaries Dramas Romantic Movies Dramas Romant
ic Movies Comedies International Movies Music & Musicals Comedies Dramas Independent Movies Comedies Dramas I
ndependent Movies Action & Adventure Dramas Action & Adventure Thrillers Documentaries Action & Adventure Dra
mas Comedies Dramas Romantic Movies Comedies Romantic Movies Comedies Dramas International Movies Cult Movies
Horror Movies Dramas Comedies Dramas Independent Movies Dramas Sports Movies Dramas Independent Movies LGBTQ
Movies Action & Adventure Dramas Sports Movies Action & Adventure Action & Adventure Action & Adventure Horro
r Movies Thrillers Horror Movies Thrillers Action & Adventure Dramas International Movies Action & Adventure
Dramas International Movies Action & Adventure Action & Adventure Action & Adventure Comedies Action & Advent
ure Classic Movies Cult Movies Children & Family Movies Action & Adventure Comedies Romantic Movies Dramas In
ternational Movies LGBTQ Movies Thrillers Thrillers Thrillers Thrillers Comedies Dramas Independent Movies Co
medies Dramas Romantic Movies Action & Adventure Classic Movies Dramas Romantic Movies Thrillers Horror Movie
s Comedies Dramas Independent Movies Thrillers Stand-Up Comedy Comedies International Movies Dramas Children
& Family Movies Comedies Children & Family Movies Dramas Faith & Spirituality Horror Movies International Mov
ies Thrillers Stand-Up Comedy Dramas Independent Movies International Movies Dramas Independent Movies Intern
ational Movies Dramas Independent Movies International Movies Dramas Independent Movies International Movies
Dramas International Movies Sports Movies Action & Adventure Dramas International Movies Documentaries Intern
ational Movies Dramas International Movies Thrillers Children & Family Movies Comedies Sci-Fi & Fantasy Telenovelas

In [63]: wordcloud = WordCloud(width = 800, height = 800,
                                background_color = 'white',
                                stopwords = None,
                                min_font_size = 10).generate(all_genres)

In [94]: plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)
plt.title('Categories Wordcloud')
plt.savefig('wordcloud')
plt.show()
```



- From the data, it is evident that international movies followed by Drama and comedy are the most popular genres in which movies are produced more.

After how many days the movie will be added to Netflix after the release of the movie ?

In [65]: `movies_data['df_date_added'] = pd.to_datetime(movies_data['date_added']).dt.year`

C:\Users\Admin\AppData\Local\Temp\ipykernel_708\2467727582.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

`movies_data['df_date_added'] = pd.to_datetime(movies_data['date_added']).dt.year`

In [66]: `movies_after_2016 = movies_data[movies_data['release_year'] > 2016]
movies_after_2016.info()`

```
<class 'pandas.core.frame.DataFrame'>  
Index: 3577 entries, 0 to 8787  
Data columns (total 15 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   show_id         3577 non-null   object  
1   type            3577 non-null   object  
2   title           3577 non-null   object  
3   director        3577 non-null   object  
4   cast            3577 non-null   object  
5   country         3577 non-null   object  
6   date_added      3577 non-null   object  
7   release_year    3577 non-null   int64  
8   rating          3577 non-null   object  
9   duration        3577 non-null   object  
10  listed_in       3577 non-null   object  
11  description     3577 non-null   object  
12  year            3577 non-null   object  
13  month           3577 non-null   object  
14  df_date_added   3577 non-null   int32  
dtypes: int32(1), int64(1), object(13)  
memory usage: 433.2+ KB
```

In [67]: `movies_after_2016['date_added_year'] = pd.to_datetime(movies_after_2016['date_added'])`

C:\Users\Admin\AppData\Local\Temp\ipykernel_708\1585031695.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

`movies_after_2016['date_added_year'] = pd.to_datetime(movies_after_2016['date_added'])`

In [68]: `movies_after_2016['release_year_dt'] = pd.to_datetime(movies_after_2016['release_year'], format='%Y')`

C:\Users\Admin\AppData\Local\Temp\ipykernel_708\1807028980.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

`movies_after_2016['release_year_dt'] = pd.to_datetime(movies_after_2016['release_year'], format='%Y')`

In [69]: `df['datetime_column'] = pd.to_datetime(df['release_year'], format='%Y')`

```
In [70]: movies_after_2016.tail()
```

Out[70]:

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | year |
|------|---------|-------|---------------------------|-----------------|---|----------------|-------------------|--------------|--------|----------|--|---|------|
| 8774 | s8775 | Movie | يوم الدين | Abu Bakr Shawky | Rady Gamal, Ahmed Abdelhafiz, Shahira Fahmy, S... | Egypt | October 22, 2019 | 2018 | TV-14 | 97 min | Dramas, Independent Movies, International Movies | A man cured of leprosy and a young orphan leav... | 2019 |
| 8774 | s8775 | Movie | يوم الدين | Abu Bakr Shawky | Rady Gamal, Ahmed Abdelhafiz, Shahira Fahmy, S... | Austria | October 22, 2019 | 2018 | TV-14 | 97 min | Dramas, Independent Movies, International Movies | A man cured of leprosy and a young orphan leav... | 2019 |
| 8774 | s8775 | Movie | يوم الدين | Abu Bakr Shawky | Rady Gamal, Ahmed Abdelhafiz, Shahira Fahmy, S... | United States | October 22, 2019 | 2018 | TV-14 | 97 min | Dramas, Independent Movies, International Movies | A man cured of leprosy and a young orphan leav... | 2019 |
| 8786 | s8787 | Movie | You Can Tutu | James Brown | Lily O'Regan, Jeannettzy Enriquez Borges, Joel... | United Kingdom | December 31, 2017 | 2017 | TV-G | 87 min | Children & Family Movies | A gifted young ballet dancer struggles to find... | 2017 |
| 8787 | s8788 | Movie | You Can't Fight Christmas | Kenny Young | Brely Evans, Andra Fuller, Persia White, Porsc... | United States | December 19, 2017 | 2017 | TV-PG | 85 min | Comedies, Romantic Movies | After stepping in to revitalize his family's h... | 2017 |

```
In [71]: movies_after_2016['year_difference']=(movies_after_2016['date_added_year']-movies_after_2016['release_year_dt']).dt.days
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_708\2308321442.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
movies_after_2016['year_difference']=(movies_after_2016['date_added_year']-movies_after_2016['release_year_dt']).dt.days
```

```
In [72]: mode_difference= movies_after_2016['year_difference'].mode()[0]
```

```
In [73]: mode_difference
```

Out[73]: 424

Observation

- The best time for Netflix to add new content depends on a combination of factors including audience behavior, market dynamics, competition, and strategic objectives.
- In this case considering past recent data (movies released after 2016) it can be seen that new movies are added on Netflix after approximately 334 days or after an year followed by a successful theatrical release.

What is the best time of the year when maximum content get added on the Netflix?

```
In [74]: df.head(3)
```

```
Out[74]:
```

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | year | |
|---|---------|---------|----------------------|------------------|---|-----------------|--------------------|--------------|--------|-----------|---|---|------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | Unknown cast | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmm... | 2021 | S |
| 1 | s2 | TV Show | Blood & Water | Unknown director | 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... | 2021 | S |
| 2 | s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi... | Unknown 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 lor... | 2021 | S |

```
In [75]: month_year= df.groupby(['year','month'])['show_id'].count().reset_index()
```

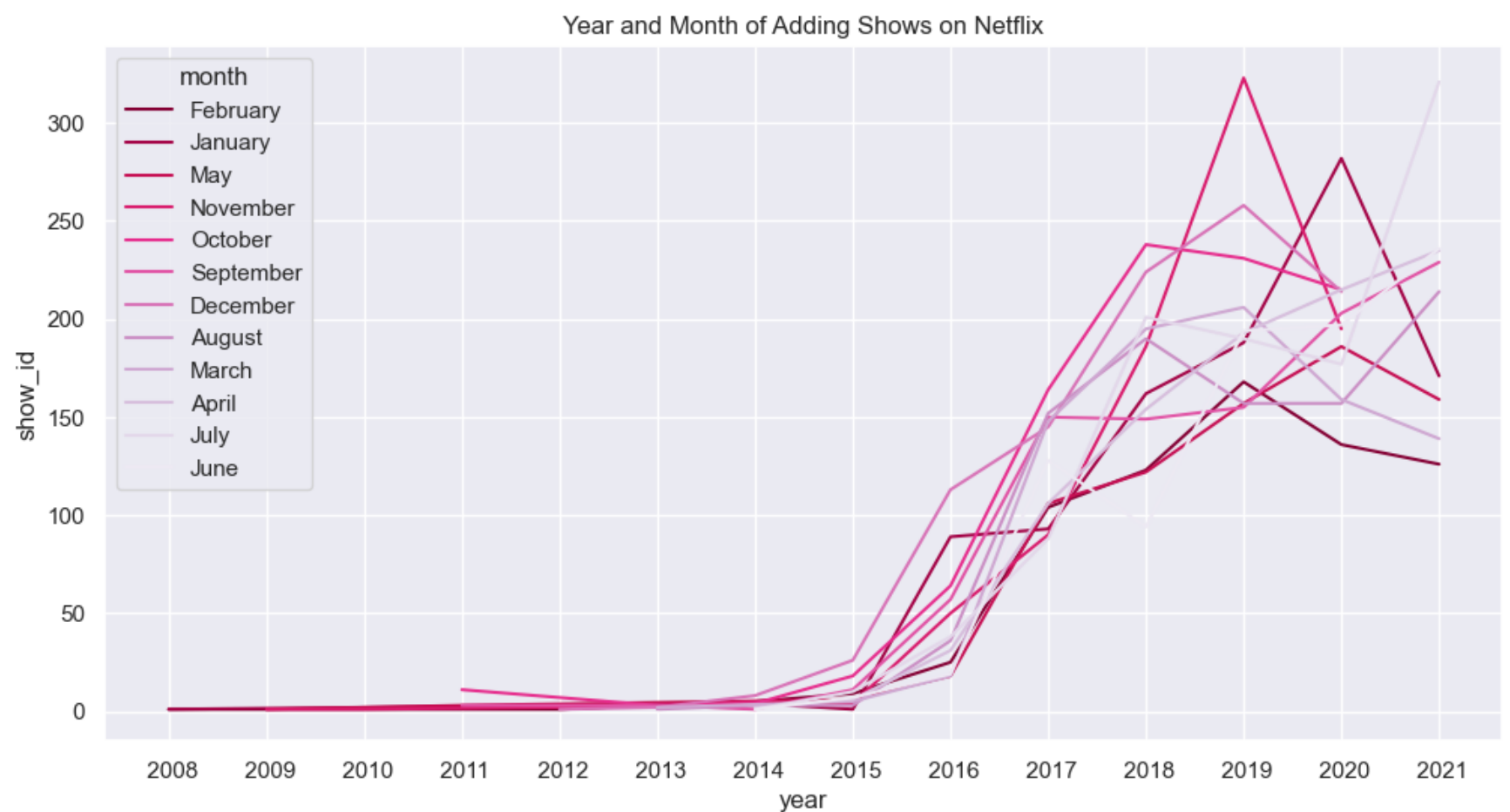
```
In [76]: month_year.sort_values(by=('year'),ascending=False)
```

```
Out[76]:
```

| | year | month | show_id |
|-----|------|-----------|---------|
| 107 | 2021 | September | 229 |
| 106 | 2021 | May | 159 |
| 105 | 2021 | March | 139 |
| 104 | 2021 | June | 236 |
| 103 | 2021 | July | 321 |
| ... | ... | ... | ... |
| 4 | 2010 | November | 1 |
| 2 | 2009 | May | 1 |
| 3 | 2009 | November | 1 |
| 1 | 2008 | January | 1 |
| 0 | 2008 | February | 1 |

108 rows × 4 columns

```
In [93]: plt.figure(figsize=(12,6))
sns.lineplot(data=month_year,x='year',y='show_id',hue='month',palette='PuRd_r')
plt.savefig('year and month of adding shows')
plt.title('Year and Month of Adding Shows on Netflix')
plt.show()
```



- The number of shows getting added is increasing with each year until 2020.
- Also, months in the last quarter of the year (Oct-Dec) have more shows being added than the other months of the year. This could be because US has its festive season in Dec and India also has Diwali in Oct-Nov

In []: ►