

Netflix-Data Exploration and Visualisation

Business Problem

Analyze the data and generate insights that could help Netflix team deciding which type of shows/movies to produce and how they can grow the business in different countries

Observations on the shape of data, data types, conversion of categorical attributes to 'category' (If required), missing value detection, statistical summary.

Importing libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Reading File

```
df=pd.read_csv('netflix.csv')
```

First 10 rows

```
df.head(10)
```

	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, K'omi Ngema, Gail Mababane, Thabani...	South Africa	September 24, 2021	2021	TV-MA	Seasons 2	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 Gotoos, Samuel Jaay, Nabli...	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 L...
5	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hannah Harshbarger, Sacha P...	NaN	September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Mysteries	The arrival of a charismatic young priest bring...
6	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, Chris Marler...	NaN	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, Oyafunmiluke Ogundimu, Alexander D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent TV Shows, International Movies	On a photo shoot in Ghana, an American model s...
8	s9	TV Show	The Great British Baking Show	Andy Devine	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021	TV-14	Seasons 9	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...

Last 10 rows

```
df.tail(10)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
8797	s8798	TV Show	Zak Storm	NaN	Mihraaj Johnston, Jessica Gee-George, Christin...	United States, France, South Korea, Indonesia	September 13, 2018	2016	TV-Y7	3 Seasons	Kids' TV	Teen surfer Zak Storm is mysteriously transpor...
8798	s8799	Movie	Zed Plus	Chandra Prakash Dewvedi	Adil Hussain, Singh, K.K. Raina, Sarjaya M...	India	December 31, 2019	2014	TV-MA	131 min	Comedies, Dramas, International Movies	A phandering small-town romantic's political...
8799	s8800	Movie	Zenda	Awadhoot Gupte	Siddharth Chaudhkar, Sacha P...	India	February 15, 2018	2009	TV-14	120 min	Dramas, International Movies	A change in the leadership of a political part...
8800	s8801	TV Show	Zindagi Gulzar Hai	NaN	Sanam Seead, Fawad Khan, Ayesha Omer, Mehru...	Pakistan	December 15, 2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...	Strong-willed, middle-aged, Karachi and cashfe...
8801	s8802	Movie	Zinzana	Majid Al-Sultan, Asah Al-Jabri...	Ali Sulman, Asah Al-Jabri...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recovering from a scandalous affair, a former sp...
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 Harrison, 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, Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubani	Mozes Singh	Vicky Kaushal, Sara-Jane Khan, Dias Raaghav, Chani...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musical	A scrappy but poor boy worms his way into a ty...

Shape of the dataset

```
df.shape
```

(8807, 12)

Different attributes in data

```
df.columns
```

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

Data types of all the attributes

```
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: category(1), int64(1), object(10)
memory usage: 825.8+ KB
```

df['type'].unique()

array(['Movie', 'TV Show'], dtype=object)

Converting into categorical attributes

```
df['type']=df['type'].astype('category')
```

Datatype of 'type' column changed

```
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    category
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: category(1), int64(1), object(10)
memory usage: 765.7+ KB
```

Missing value detection

```
df.isnull().sum()
```

```
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
```

Statistical summary of 'int' data type

```
df.describe(include='int64')
```

release_year

```
count    8807.000000
mean     2014.180198
std       8.819312
min       1925.000000
25%      2013.000000
50%      2017.000000
75%      2019.000000
max       2021.000000
```

Statistical summary of 'object' data type

```
df.describe(include='object').T
```

count

```
show_id    8807    8807
title      8807    8807
director   6173    4528
cast       7982    7692
country    7976    748
date_added 8797    1767
rating     8803    17
duration   8804    220
listed_in  8807    514
description 8807    8775
```

freq

```
show_id    1
title      1
director   19
cast       19
country    2818
date_added 309
rating     17
duration   173
listed_in  362
description 4
```

Unnesting date_added attribute

```
date = df[['date_added']].dropna()
```

```
date['year'] = date['date_added'].apply(lambda x: x.split('.')[0])
date['month'] = date['date_added'].apply(lambda x: x.split('.')[1])
```

date['year'].head()

```
0    2021
1    2021
2    2021
3    2021
4    2021
Name: year, dtype: object
```

date['month'].head()

```
0    September
1    September
2    September
3    September
4    September
Name: month, dtype: object
```

Non-Graphical Analysis: Value counts and unique attributes.

Unique values of 'type' attribute

```
df['type'].unique()
```

['Movie', 'TV Show']

Categories (2, object): ['Movie', 'TV Show']

Total count of unique values in 'type' attribute

```
df['type'].value_counts()
```

```
Movie      6131
TV Show   2676
Name: type, dtype: int64
```

Insights :Total number of movies is greater than TV shows

Unique values of month which movies/tv shows are added

```
date['month'].unique()
```

```
array(['September', 'January', 'July', 'June', 'May', 'April', 'March',
       'February', 'August', 'December', 'November', 'October'],
      dtype=object)
```

date['month'].value_counts()

```
July          827
December     813
September    770
April        764
October      760
August       755
March        742
January      738
June         728
November     705
May          632
February     563
Name: month, dtype: int64
```

Insights :Most of movies and TV shows are added on Netflix in July

Unique values of year which movies/tv shows are added

```
date['year'].unique()
```

```
array(['2021', '2020', '2019', '2018', '2017', '2016', '2015', '2014',
       '2013', '2012', '2011', '2009', '2008', '2010'], dtype=object)
```

date['year'].value_counts()

```
2019    2016
2020    1879
2018    1649
2021    1498
2017    1188
2016    429
2015    62
2014    24
2011    13
2013    11
2012    3
2009    2
2008    2
2010    1
Name: year, dtype: int64
```

Insights :Most of movies and TV shows are added on Netflix in 2019

Unique values of 'release_year' attribute

```
df['release_year'].unique()
```

```
array([2020, 2021, 1993, 1996, 1998, 1999, 2010, 2013, 2017, 1975,
       1978, 1983, 1987, 2012, 2001, 2014, 2002, 2003, 2004, 2011, 2008,
       2009, 2007, 2005, 2006, 1994, 2015, 2019, 2016, 1982, 1989, 1990,
       1991, 1999, 1986, 1992, 1984, 1980, 1961, 2000, 1995, 1985, 1976,
       1999, 1988, 1981, 1972, 1964, 1985, 1954, 1979, 1958, 1956, 1963,
       1970, 1973, 1925, 1974, 1960, 1966, 1971, 1962, 1969, 1977, 1967,
       1968, 1965, 1946, 1942, 1955, 1944, 1947, 1943], dtype=int64)
```

Total count of unique values in 'release_year' attribute

```
df['release_year'].value_counts()
```

```
2018    1147
2017    1032
2019    1030
2020    953
2016    902
```

1959 1
1925 1
1961 1
1947 1
1966 1
Name: release_year, Length: 74, dtype: int64

Insights :Most number of movies and TV shows are released in 2018

Unique values of 'rating' attribute

```
df['rating'].unique()
```

```
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', nan,
       'TV-Y7-FV', 'UR'], dtype=object)
```

Total count of unique values in 'rating' attribute

```
df['rating'].value_counts().head()
```

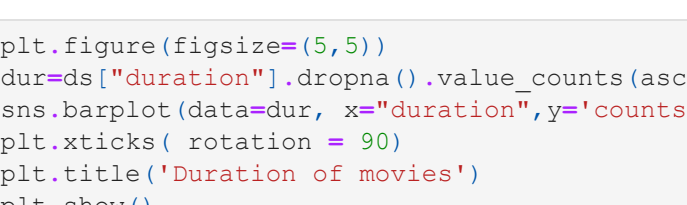
```
TV-MA    3207
TV-14    2160
TV-PG    863
R         759
PG-13    490
Name: rating, dtype: int64
```

Insights :Most number of movies and TV shows are rated as TV-MA

Visual Analysis - Univariate, Bivariate after pre-processing of the data

Comparing movies and TV shows on Netflix

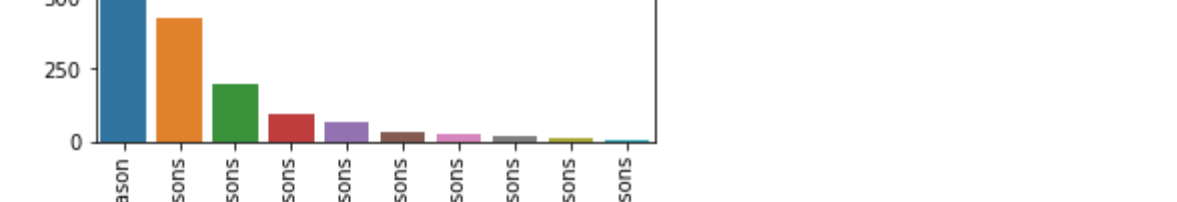
```
plt.figure(figsize=(5,5))
type=df['type'].dropna().value_counts().rename_axis('type').reset_index(name='counts')
sns.barplot(data=type, x='type', y='counts')
plt.title('Comparison of movies / TV shows')
plt.show()
```



Insights :Total number of movies is more than TV shows

Trend of content produced on Netflix over the last 22 years

```
plt.figure(figsize=(10,8))
data_movies=df[df['type']=='Movie']
x_movies=list(data_movies.groupby(['release_year'])['title'].nunique().tail(22).index)
y_movies=list(data_movies.groupby(['release_year'])['title'].nunique().tail(22))
sns.lineplot(x=x_movies,y=y_movies,label='Movie')
data_shows=df[df['type']=='TV Show']
x_shows=list(data_shows.groupby(['release_year'])['title'].nunique().tail(22).index)
y_shows=list(data_shows.groupby(['release_year'])['title'].nunique().tail(22))
sns.lineplot(x=x_shows,y=y_shows,label='TV Show')
plt.xticks(rotation = 90)
plt.xlabel('release_year')
plt.ylabel('release count')
plt.title('Trend of content on Netflix')
plt.legend()
plt.show()
```

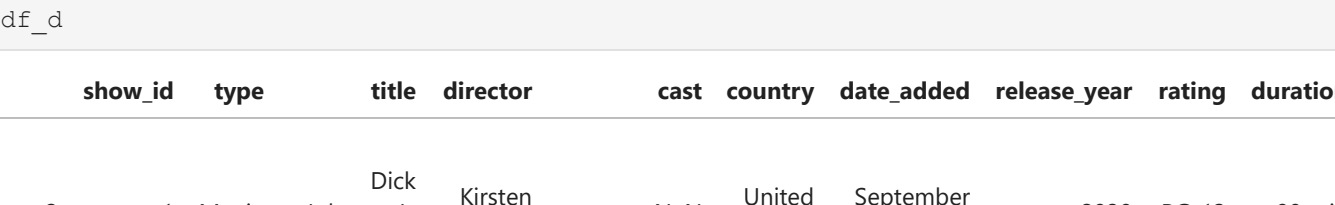


Insights :Till 2018 the trend of movies was higher than that of TV shows,but after 2018 movies is showing decrease in trend.In 2021 TV shows trend is higher than that of movies.

Recommendation :Netflix should produce more TV shows than movies as per the trend.

Release Year of Movies/TV shows from 2000

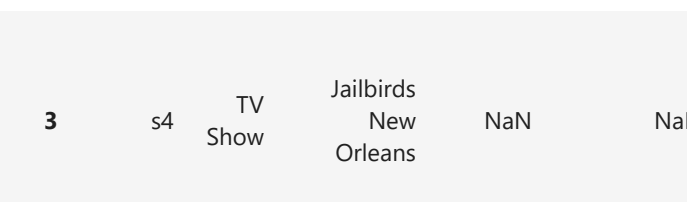
```
plt.figure(figsize=(5,5))
year=df['release_year'].astype('Int64')
gh = sns.histplot(x = year,bins=np.arange(2000, 2022, step=2))
sns.barplot(data=year, x='release_year', y='counts')
plt.xticks( rotation = 90)
plt.xlabel('Release year')
for i in gh.patches:
    gh.annotate(i.get_height(),format(i.get_height()*(i.get_x()+2.5,i.get_height()+50) )
plt.xticks(rotation=2000, 2022, step=2)
plt.show()
```



Insights :Most of Movie/TV shows are released between the year 2018 and 2020

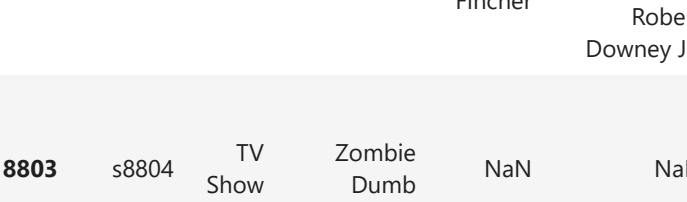
Top 5 years when most number of movies / TV shows are released

```
plt.figure(figsize=(5,5))
year=df['release_year'].dropna().value_counts().rename_axis('release_year').reset_index(name='counts')[:5]
sns.barplot(data=year, x='release_year', y='counts')
plt.xticks( rotation = 90)
plt.xlabel('Top 5 years most number of movies / TV shows are released')
plt.show()
```



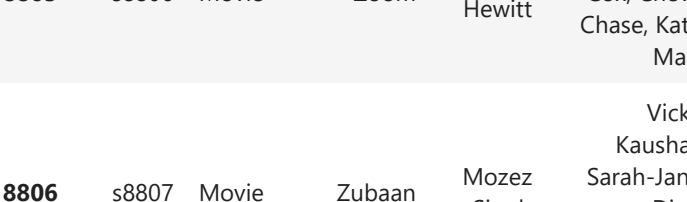
Insights :Most number of movies and TV shows are released in 2018

```
plt.figure(figsize=(5,10))
year=df['release_year']
sns.boxplot(data=year)
plt.xlabel('Movies / TV shows')
plt.ylabel('release_year')
plt.xticks( rotation = 90)
plt.title('Movies / TV shows release year')
plt.show()
```



Rating of movies and TV shows on Netflix

```
plt.figure(figsize=(5,5))
rate=df['rating'].replace({'74 min':np.nan,'84 min':np.nan,'66 min':np.nan}).dropna()
sns.barplot(data=rate, x='duration', y='counts')
plt.xticks( rotation = 90)
plt.title('Rating of movies / TV shows on Netflix')
plt.show()
```



Insights :Most number of movies and TV shows on Netflix are rated as TV-MA

Recommendation :Netflix should produce more TV-MA rated movies and TV shows.

Duration of movies

```
df[df['type']=='Movie']
```

```
plt.figure(figsize=(10,5))
dur=df['duration'].dropna().value_counts(ascending=False).rename_axis('duration').reset_index(name='counts')[:1]
sns.barplot(data=dur, x='duration', y='counts')
plt.xticks( rotation = 90)
plt.title('Duration of movies')
plt.show()
```



Insights :Most of movies have 90 minutes duration

Recommendation :Netflix should produce movies having 90 minutes duration.

Duration of TV shows

```
df[df['type']=='TV Show']
```

```
plt.figure(figsize=(5,5))
dur=df['duration'].dropna().value_counts(ascending=False).rename_axis('duration').reset_index(name='counts')[:1]
sns.barplot(data=dur, x='duration', y='counts')
plt.xticks( rotation = 90)
plt.title('Duration of movies')
plt.show()
```


Insights :Most number of TV shows have 1 season duration.

Recommendation :Netflix should produce TV shows having 1 season duration.

Pre-processing of the data (unnesting of the data in columns like director,country,listed_in)

```
df=df.copy()
```

Unnesting 'director' column

```
df=df.assign(directors=df['director'].str.split(",").explode("director")
```

```
df['director']=df['director'].str.strip()
```

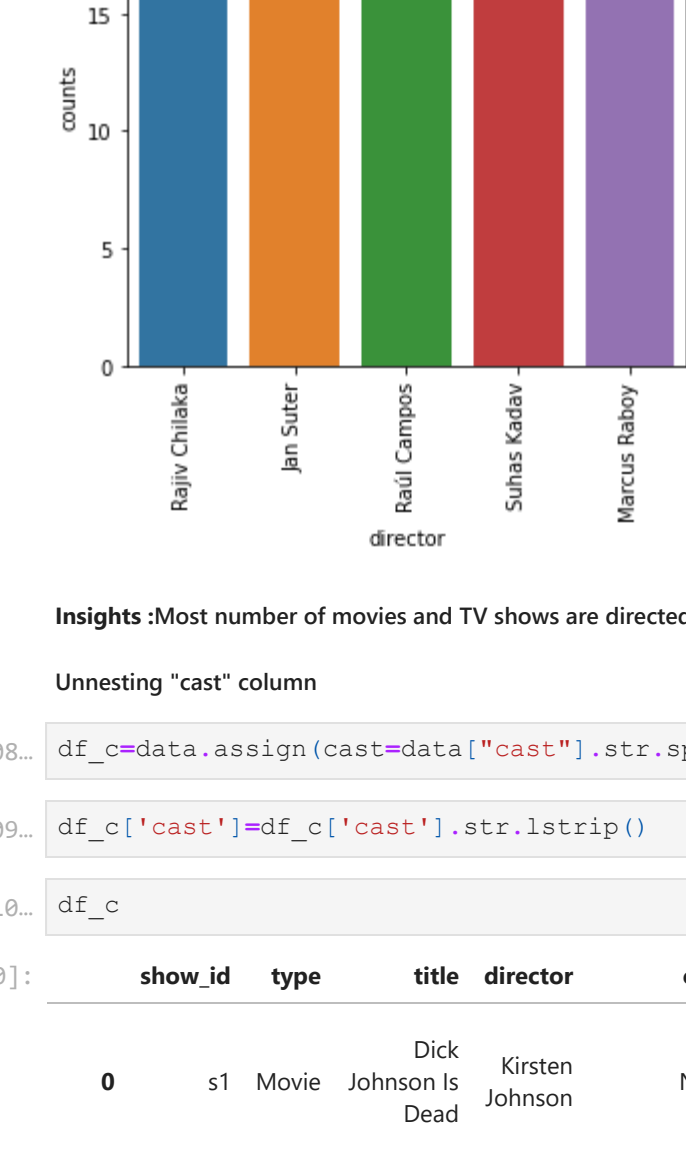
```
df['director'].dropna(inplace=True)
```

```
df['director']
```

	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, K'omi Ngema, Gail Mababane, Thabani...	South Africa	September 24, 2021	2021	TV-MA	Seasons 2	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3											


```
In [207]: df_c=df_c[df_c['director'].value_counts().rename_axis('director').reset_index(name='counts')[:5]]
sns.barplot(data=df_c, x='director', y='counts')
plt.xticks(rotation=90)
plt.title('Top 5 directors with most number of movies / TV shows')
plt.show()
```

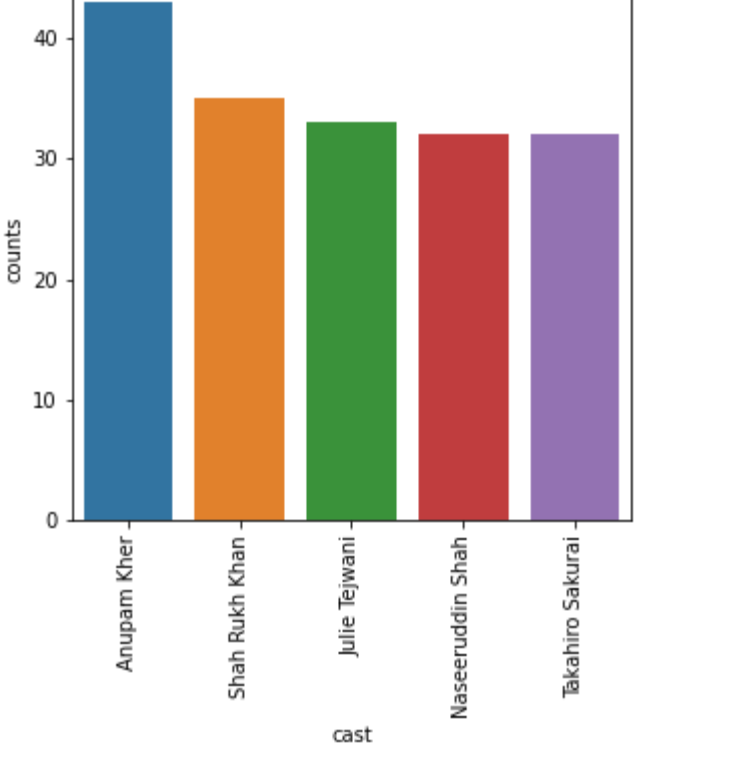
Top 5 directors with most number of movies / TV shows



Insights: Most number of movies and TV shows are directed by Rajiv Chhaya

```
df_c=df_c.assign(cast=df_c['cast'].str.split(",").explode("cast"))
df_c['cast']=df_c['cast'].str.strip()
```

df_c

1	s2	TV Show	Blood & Water	NaN	Khosi Ngema	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...
1	s2	TV Show	Blood & Water	NaN	Gail Mababane	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...
1	s2	TV Show	Blood & Water	NaN	Thabang Molaba	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...
...
8806	s8807	Movie	Zubaneh Moze Singh	Marish Chaudhary	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...
8806	s8807	Movie	Zubaneh Moze Singh	Meghna Malik	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...
8806	s8807	Movie	Zubaneh Moze Singh	Mallikarjun Rauti	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...
8806	s8807	Movie	Zubaneh Moze Singh	Anita Shabash	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...
8806	s8807	Movie	Zubaneh Moze Singh	Chittaranjan Tipthary	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...
64951 rows × 12 columns												
<pre>df_c.shape (64951, 12)</pre>												
Top 5 actors with most number of movies / TV shows												
<pre>df_c['cast'].value_counts().head()</pre>												
<pre>Anupam Kher 43 Shah Rukh Khan 35 Julie Teysawari 33 Naseeruddin Shah 32 Takahiro Sakurai 32 Name: cast, dtype: int64</pre>												
<pre>plt.figure(figsize=(5,5)) cat=df_c['cast'].value_counts().rename_axis('cast').reset_index(name='counts')[5] sns.barplot(data=cat,x='cast',y='counts') plt.xticks(rotation=90) plt.title('Top 5 directors with most number of movies / TV shows') plt.show()</pre>												
Top 5 directors with most number of movies / TV shows												
												
Insights :Most number of movies and TV shows are acted by Anupam Kher												
Unnesting 'country' column												
<pre>df_c=df_c.assign(country=df_c['country'].str.split(",").explode("country"))</pre>												
<pre>df_c['country']=df_c['country'].str.strip()</pre>												
<pre>df_c</pre>												
	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...
Amra Qamata, Khosi Ngema, Gail Mababane, Thabang Molaba												
1	s2	TV Show	Blood & Water	NaN	Khosi Ngema, Gail Mababane, Thabang Molaba	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...

64951 rows x 12 columns

```
In [211]: df_c.shape
Out[211]: (64951, 12)
```

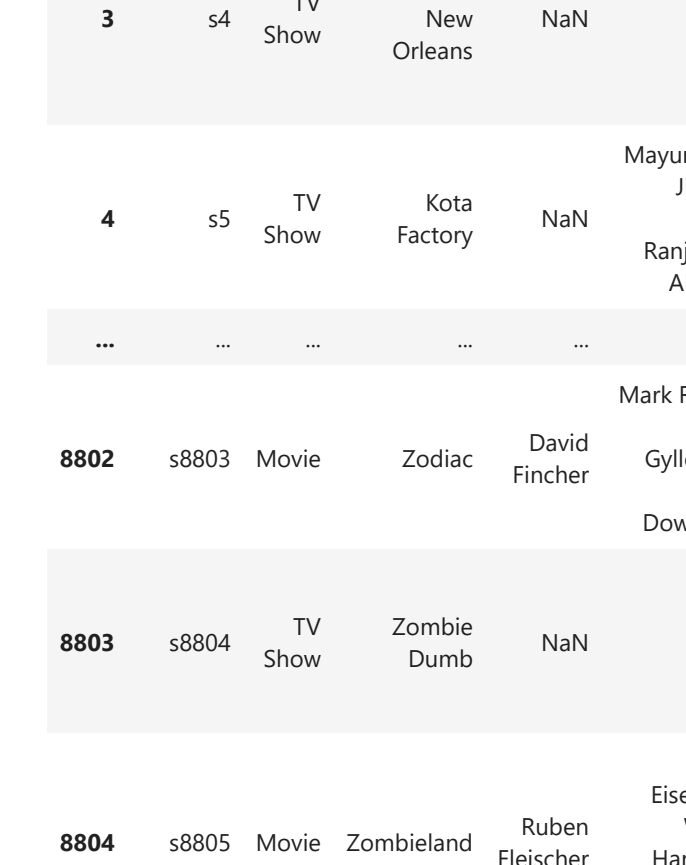
Top 5 actors with most number of movies / TV shows

```
In [212]: df_c['cast'].value_counts().head()
```

```
Anupam Kher 43
Shah Rukh Khan 33
Julie Tejlani 33
Naseeruddin Shah 32
Takshir Sakurai 32
Name: cast, dtype: int64
```

```
In [213]: plt.figure(figsize=(5,5))
sns.barplot(data=df_c, x='cast', y='counts')
plt.xticks(rotation=90)
plt.title('Top 5 directors with most number of movies / TV shows')
plt.show()
```

Top 5 directors with most number of movies / TV shows



Insights: Most number of movies and TV shows are acted by Anupam Kher

```
df_c=df_c.assign(country=df_c['country'].str.split(",").explode("country"))
df_c['country']=df_c['country'].str.strip()
```

df_c

8805	s8806	Movie	Zoom	Peter Hewitt	Courteney Cox, Chevy Chase, Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	A so...
8806	s8807	Movie	Zubaneh Moze	Mozes Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chhann...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	but boy w...

10850 rows × 12 columns

```
df_cn.shape
(10850, 12)

Top 5 countries with most number of movies / TV shows

df_cn['country'].value_counts().head()

United States    3690
India             1046
United Kingdom   806
Canada           445
France           393
Name: country, dtype: int64

plt.figure(figsize=(5,5))
count_df_cn['country'].value_counts().rename_axis('country').reset_index(name='counts')[[:5]]
sns.barplot(data=count_df_cn, x='country', y='counts')
plt.xticks(rotation = 90)
plt.title('Top 5 countries with most number of movies / TV shows')
plt.show()
```

Top 5 countries with most number of movies / TV shows

country	counts
United States	3690
India	1046
United Kingdom	806
Canada	445
France	393

Insights: Most number of movies and TV shows are released in United States

Unnesting "listed_in" column

```
df_l = data.assign(listed_in = data["listed_in"].str.split(",").explode("listed_in"))

df_l["listed_in"] = df_l["listed_in"].str.lstrip()

df_l
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descri
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	A father's th...

Amma Qamata,

10850 rows x 12 columns

```
In [217]: df_c.shape
Out[217]: (10850, 12)
```

Top 5 countries with most number of movies / TV shows

```
In [218]: df_c['country'].value_counts().head()
```

```
United States 3690
India 1046
United Kingdom 806
Canada 445
France 393
Name: country, dtype: int64
```

```
In [219]: plt.figure(figsize=(5,5))
sns.barplot(data=df_c, x='country', y='counts')
plt.xticks(rotation=90)
plt.title('Top 5 countries with most number of movies / TV shows')
plt.show()
```

Top 5 countries with most number of movies / TV shows



Insights: Most number of movies and TV shows are released in United States

```
df_c=df_c.assign(listed_in=df_c['listed_in'].str.split(",").explode("listed_in"))
df_c['listed_in']=df_c['listed_in'].str.strip()
```

df_c

Chan...

his wo...

19323 rows × 12 columns

```
df_1.shape
(19323, 12)
```

Top 5 genre of most movies / TV shows

```
df_1["listed_in"].value_counts().head()
International Movies    2752
Dramas                 2423
Comedies               1674
International TV Shows 1351
Documentaries          869
Name: listed_in, dtype: int64
```

```
plt.figure(figsize=(5,5))
lst=df_1["listed_in"].value_counts().rename_axis('listed_in').reset_index(name='counts')[0:5]
sns.barplot(data=lst, x='listed_in', y='counts')
plt.xticks(rotation = 90)
plt.title('Top 5 genre of most movies / TV shows')
plt.show()
```

Top 5 genre of most movies / TV shows

Genre	Count
International Movies	2752
Dramas	2423
Comedies	1674
International TV Shows	1351
Documentaries	869

19323 rows x 12 columns

```
In [221]: df_c.shape
Out[221]: (19323, 12)
```

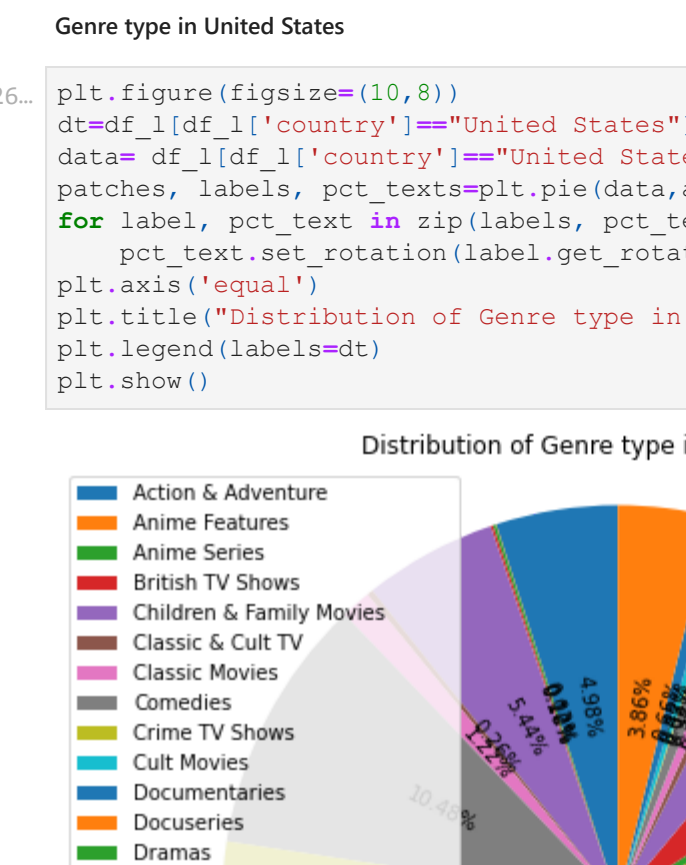
Top 5 genre of most movies / TV shows

```
In [222]: df_c['listed_in'].value_counts().head()
```

```
International Movies 2752
Dramas 2427
Comedies 1674
International TV Shows 1351
Documentaries 869
Name: listed_in, dtype: int64
```

```
In [223]: plt.figure(figsize=(5,5))
sns.barplot(data=df_c, x='listed_in', y='counts')
plt.xticks(rotation=90)
plt.title('Top 5 genre of most movies / TV shows')
plt.show()
```

Top 5 genre of most movies / TV shows



Insights: International Movies are the most among listed_in

```
In [226]: df_c[df_c['country']=='United States'].groupby(['listed_in'])['title'].unique().index
data=df_c[df_c['country']=='United States'].groupby(['listed_in'])['title'].unique()
patches, labels, pot_text=plt.pie(data,autopct='%1.2f%%', startangle=90, rotatelabels=True, pctdist=
for label, pot_text in zip(labels, pot_text):
plt.axis('equal')
plt.title('Distribution of Genre type in United States')
plt.legend(labels=dt)
plt.show()
```

Distribution of Genre type in United States

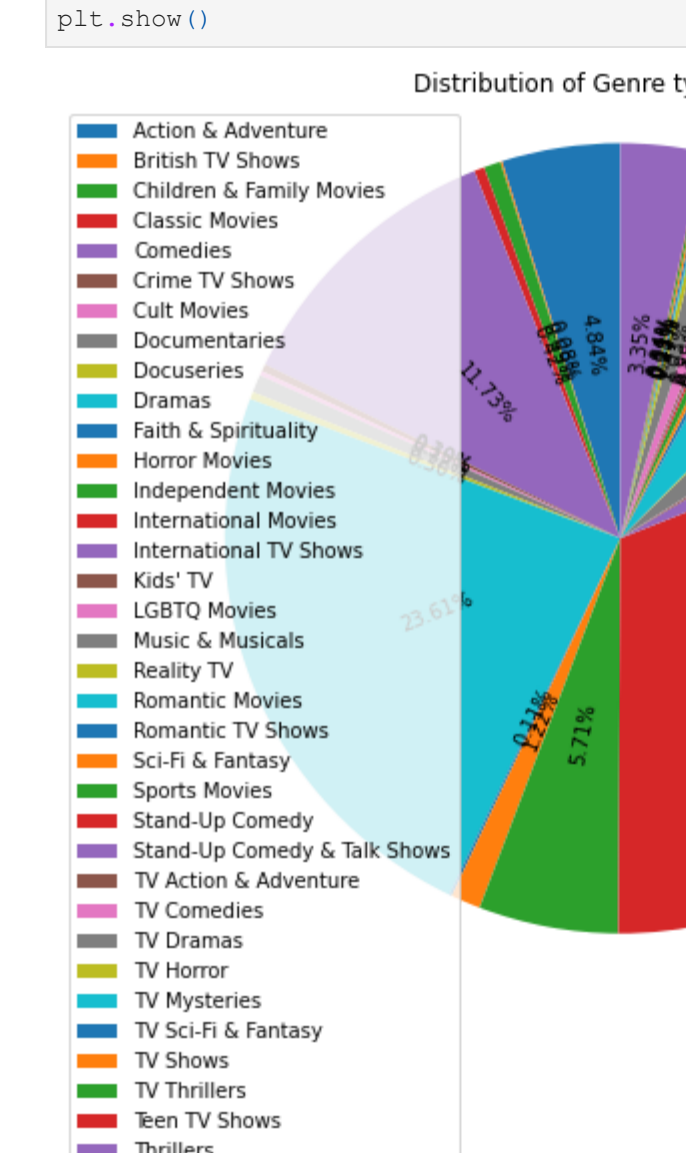


Insights: Dramas type of genre is most watched in United States

Recommendation :Netflix should produce more Dramas in United States.

```
In [227]: plt.figure(figsize=(10,8))
data=df_c[df_c['country']=='India'].groupby(['listed_in'])['title'].unique().index
data=df_c[df_c['country']=='India'].groupby(['listed_in'])['title'].unique()
patches, labels, pot_text=plt.pie(data,autopct='%1.2f%%', startangle=90, rotatelabels=True, pctdist=
for label, pot_text in zip(labels, pot_text):
plt.axis('equal')
plt.title('Distribution of Genre type in India')
plt.legend(labels=dt)
plt.show()
```

Distribution of Genre type in India

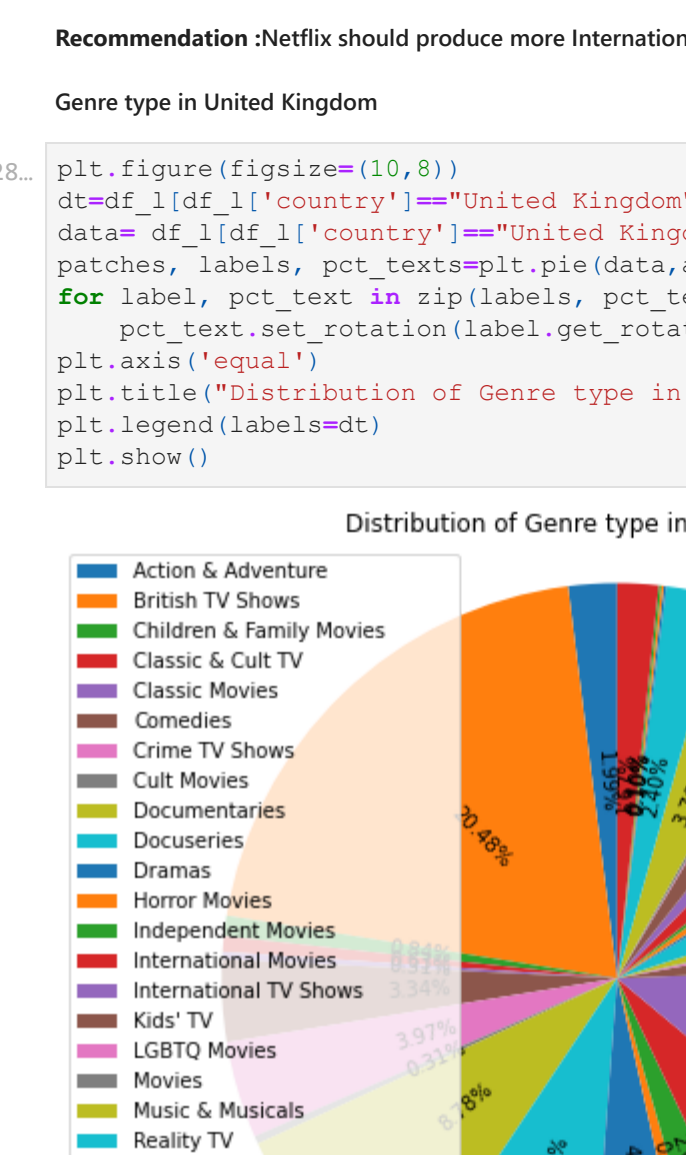


Insights: International movies type of genre is most watched in India

Recommendation :Netflix should produce more International movies in India.

```
In [228]: plt.figure(figsize=(10,8))
data=df_c[df_c['country']=='United Kingdom'].groupby(['listed_in'])['title'].unique().index
data=df_c[df_c['country']=='United Kingdom'].groupby(['listed_in'])['title'].unique()
patches, labels, pot_text=plt.pie(data,autopct='%1.2f%%', startangle=90, rotatelabels=True, pctdist=
for label, pot_text in zip(labels, pot_text):
plt.axis('equal')
plt.title('Distribution of Genre type in United Kingdom')
plt.legend(labels=dt)
plt.show()
```

Distribution of Genre type in United Kingdom

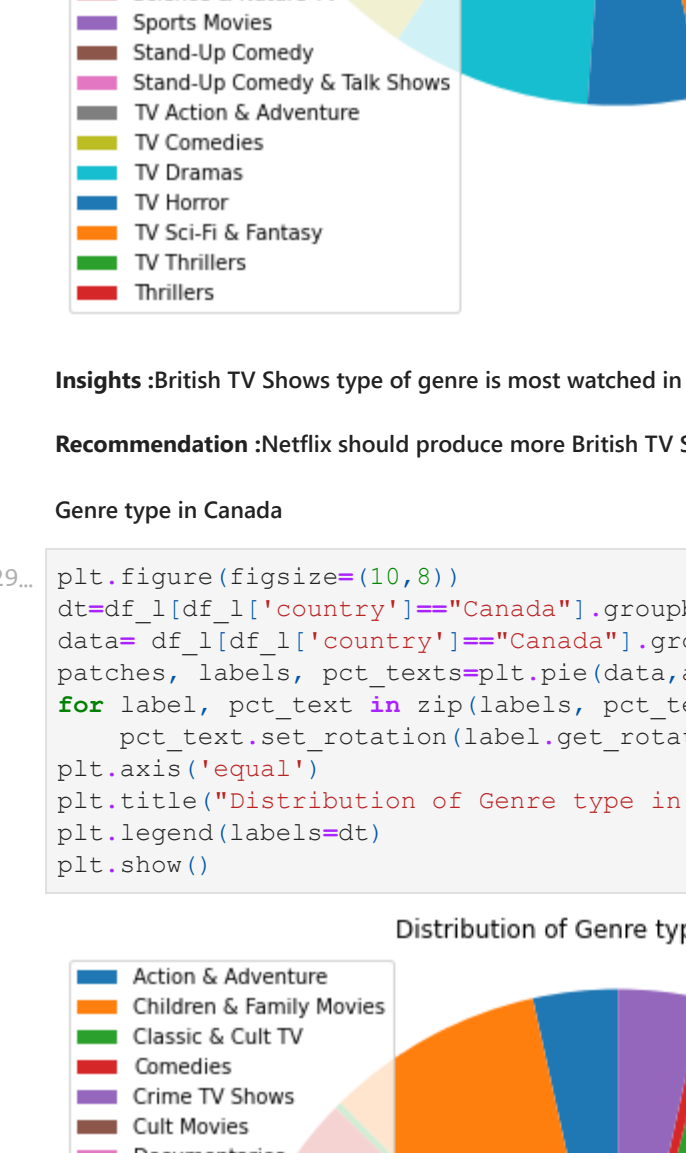


Insights: British TV Shows type of genre is most watched in United Kingdom

Recommendation :Netflix should produce more British TV Shows in United Kingdom.

```
In [229]: plt.figure(figsize=(10,8))
data=df_c[df_c['country']=='Canada'].groupby(['listed_in'])['title'].unique().index
data=df_c[df_c['country']=='Canada'].groupby(['listed_in'])['title'].unique()
patches, labels, pot_text=plt.pie(data,autopct='%1.2f%%', startangle=90, rotatelabels=True, pctdist=
for label, pot_text in zip(labels, pot_text):
plt.axis('equal')
plt.title('Distribution of Genre type in Canada')
plt.legend(labels=dt)
plt.show()
```

Distribution of Genre type in Canada

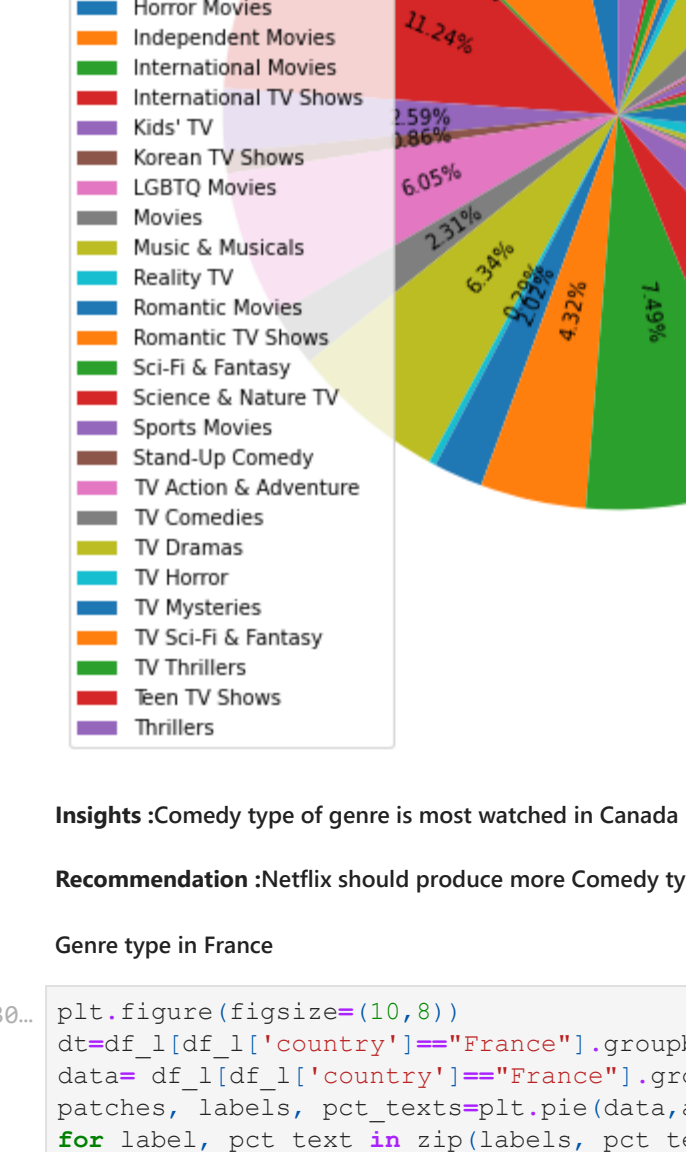


Insights: Comedy type of genre is most watched in Canada

Recommendation :Netflix should produce more Comedy type of genre in Canada.

```
In [230]: plt.figure(figsize=(10,8))
data=df_c[df_c['country']=='France'].groupby(['listed_in'])['title'].unique().index
data=df_c[df_c['country']=='France'].groupby(['listed_in'])['title'].unique()
patches, labels, pot_text=plt.pie(data,autopct='%1.2f%%', startangle=90, rotatelabels=True, pctdist=
for label, pot_text in zip(labels, pot_text):
plt.axis('equal')
plt.title('Distribution of Genre type in France')
plt.legend(labels=dt)
plt.show()
```

Distribution of Genre type in France

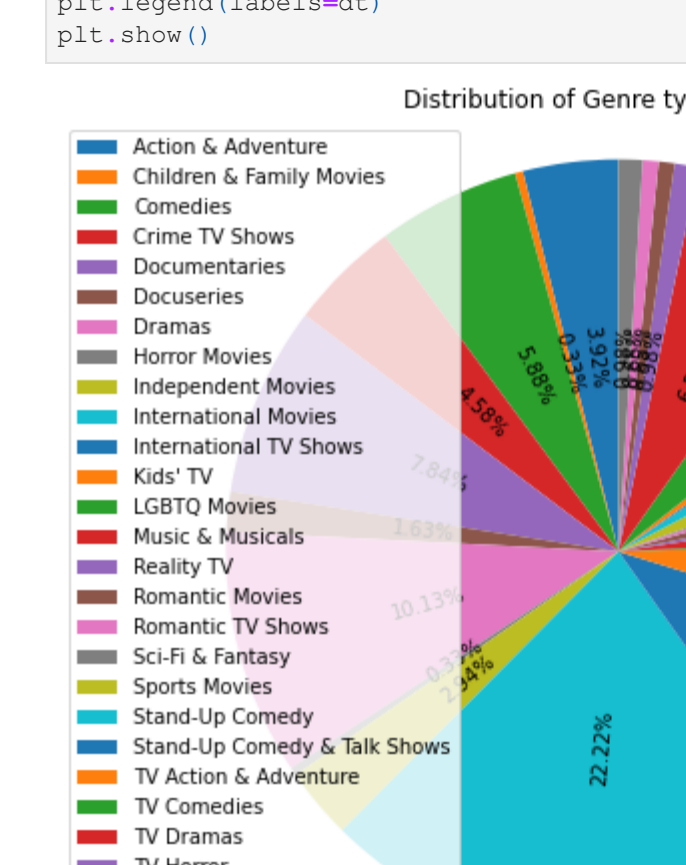


Insights: international movies type of genre is most watched in France.

Recommendation :Netflix should produce more International movies in France.

```
In [231]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='United States') & (df_c['cast']!='nan') & (df_c['type']=='Movie')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
sns.barplot(x=data_x, y=data_y)
plt.xticks(rotation=90)
plt.title('Top Successful Actor in United States for Movies')
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
gh.annotate('%.0f%%' % format(i.get_height()/i.get_x()*0.30, i.get_height()*0.2))
plt.show()
```

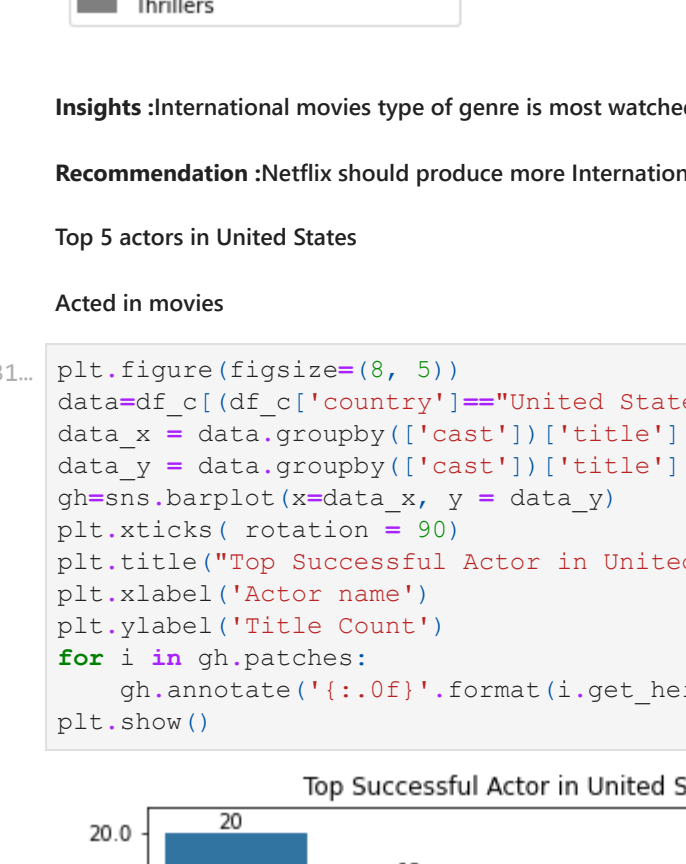
Top Successful Actor in United States for Movies



Insights in United States top 5 movie actors are Adam Sandler,Samuel L Jackson,Laura Bailey,Molly Shannon and Nicolas Cage

```
In [232]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='United States') & (df_c['cast']!='nan') & (df_c['type']=='TV Show')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
sns.barplot(x=data_x, y=data_y)
plt.xticks(rotation=90)
plt.title('Top Successful Actor in United States for TV Show')
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
gh.annotate('%.0f%%' % format(i.get_height()/i.get_x()*0.30, i.get_height()*0.3))
plt.show()
```

Top Successful Actor in United States for TV Show

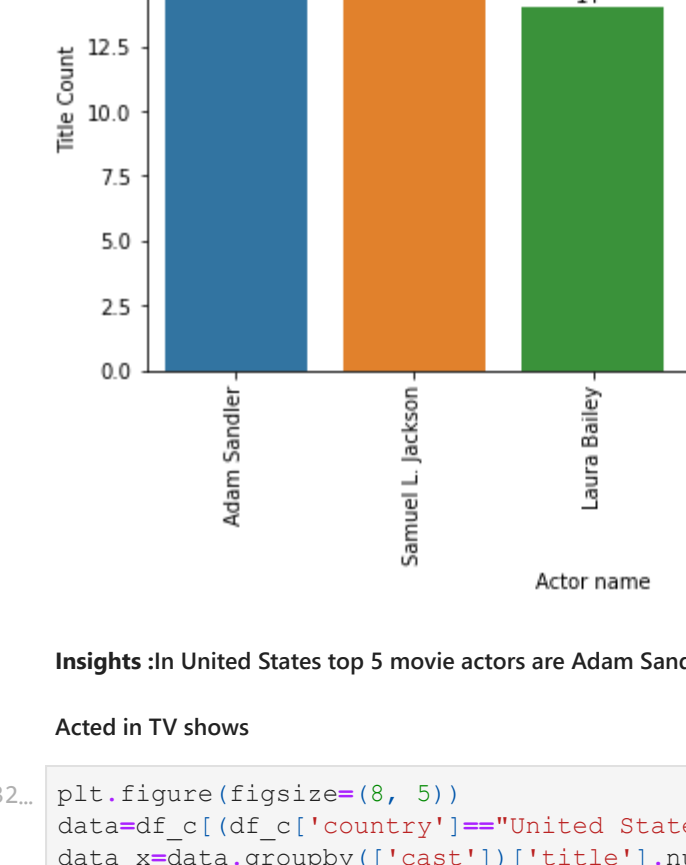


Insights in United States top 5 TV shows actors are Mike Colter,Greg Griffin,Kevin Michael,Steven Yuen and Karl Walilgren

Recommendation in United States,Netflix should produce more movies acted by Adam Sandler,Samuel L Jackson,Laura Bailey,Molly Shannon and Nicolas Cage and more TV Shows acted by Mike Colter,Greg Griffin,Kevin Michael,Steven Yuen and Karl Walilgren

```
In [233]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='India') & (df_c['cast']!='nan') & (df_c['type']=='Movie')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
sns.barplot(x=data_x, y=data_y)
plt.xticks(rotation=90)
plt.title('Top Successful Actor in India for Movies')
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
gh.annotate('%.0f%%' % format(i.get_height()/i.get_x()*0.30, i.get_height()*0.2))
plt.show()
```

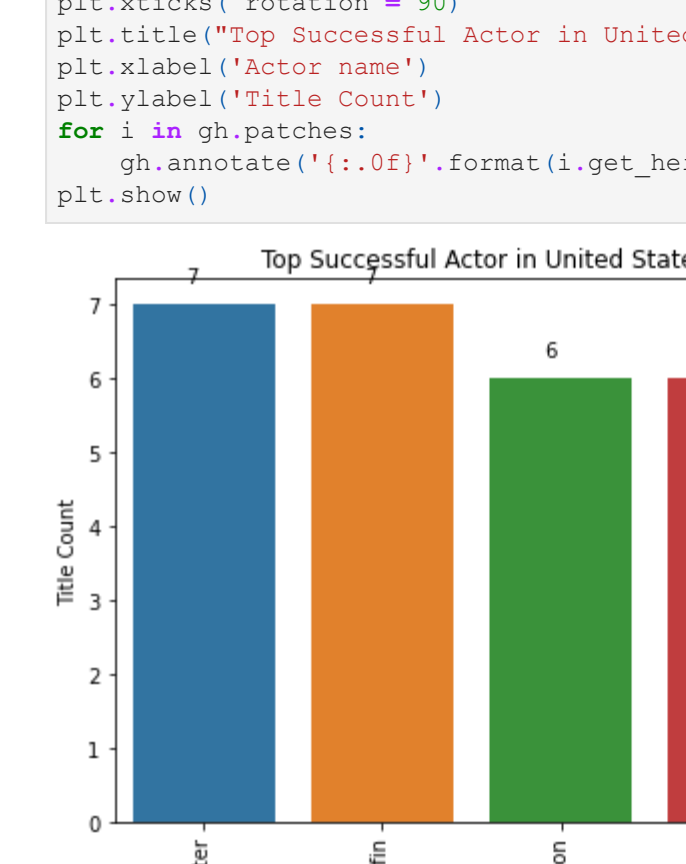
Top Successful Actor in India for Movies



Insights in India top 5 movie actors are Anupam Kher,Shah Rukh Khan,Naseeruddin Shah,Amitabh Bachchan and Akshay Kumar

```
In [234]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='India') & (df_c['cast']!='nan') & (df_c['type']=='TV Show')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
sns.barplot(x=data_x, y=data_y)
plt.xticks(rotation=90)
plt.title('Top Successful Actor in India for TV Show')
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
gh.annotate('%.0f%%' % format(i.get_height()/i.get_x()*0.30, i.get_height()*0.3))
plt.show()
```

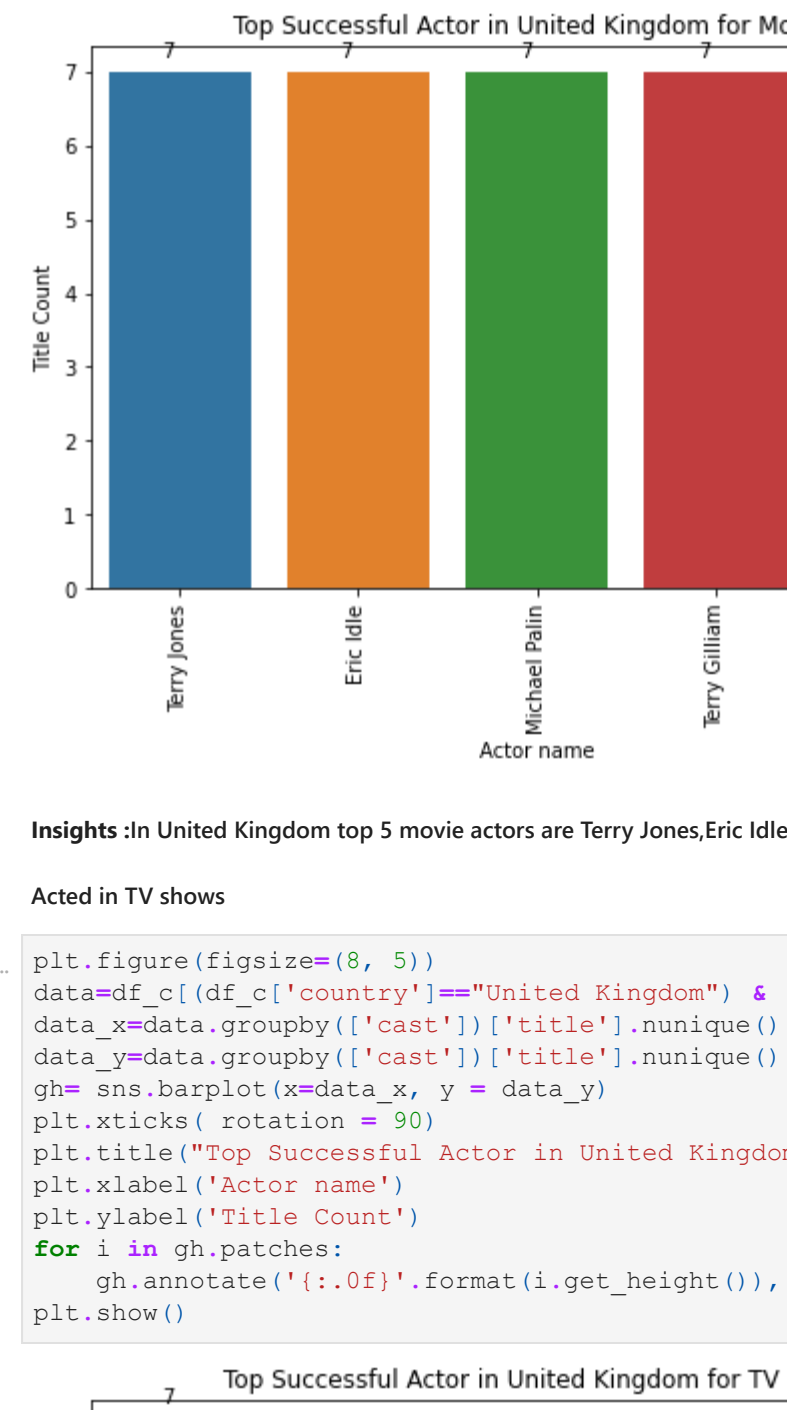
Top Successful Actor in India for TV Show



Insights in India top 5 TV show actors are Nishka Rahga,Rajesh Kava,Prakash Raj,Mona Ambegankar and Vineet Kumar

Recommendation in India,Netflix should produce more movies acted by Anupam Kher,Shah Rukh Khan,Naseeruddin Shah,Amitabh Bachchan and Akshay Kumar and more TV Shows acted by Nishka Rahga,Rajesh Kava,Prakash Raj,Mona Ambegankar and Vineet Kumar

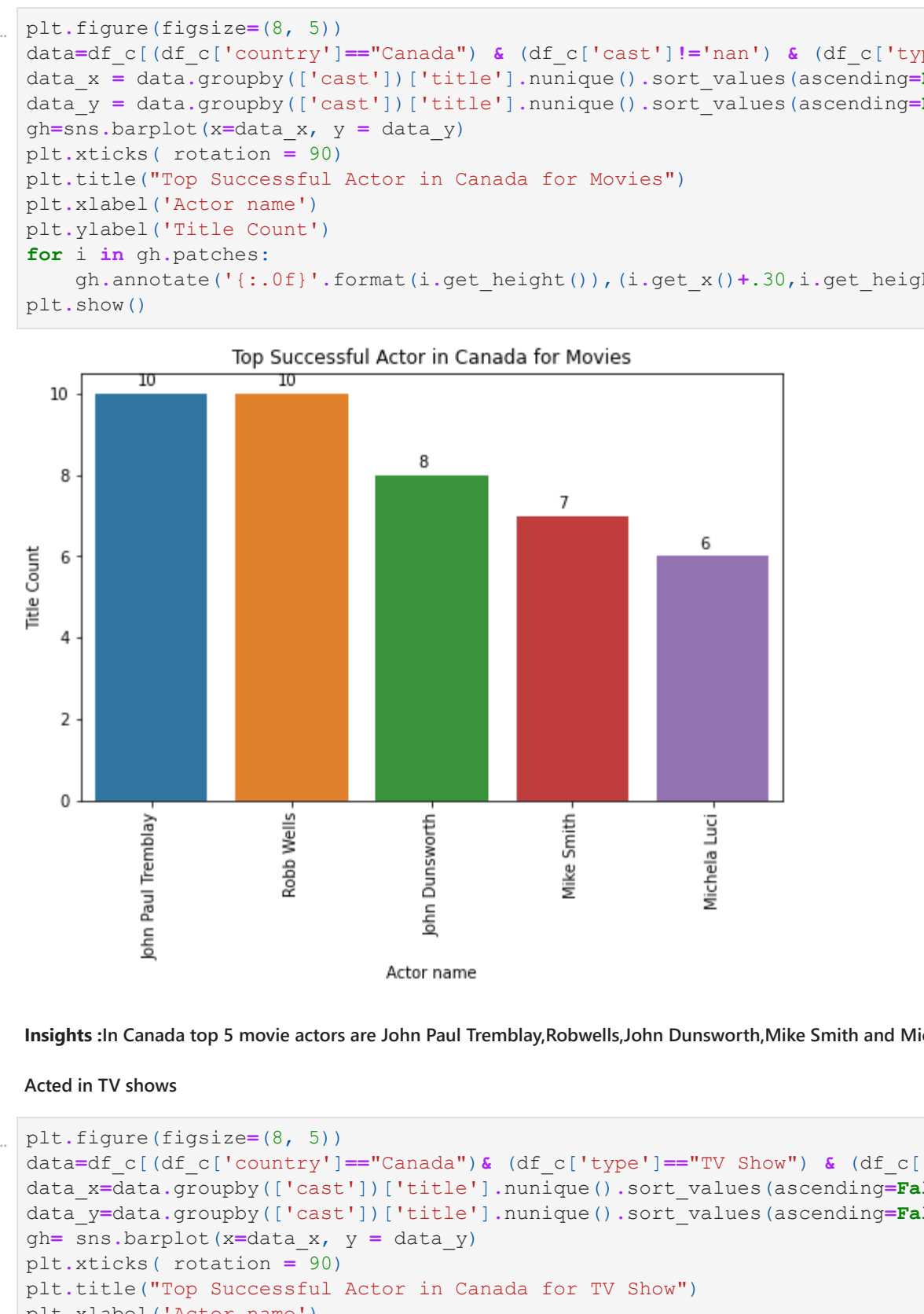
```
In [235]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='United Kingdom') & (df_c['cast']!='nan') & (df_c['type']=='Movie')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head(5).index
sns.barplot(x=data_x, y=data_y)
plt.xticks(rotation=90)
plt.title('Top Successful Actor in United Kingdom for Movies')
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
gh.annotate('%.0f%%' % format(i.get_height()/i.get_x()*0.30, i.get_height()*0.2))
plt.show()
```

Insights in United Kingdom top 5 movie actors are Terry Jones,Eric Idle,Michael Palin,Terry Gilliam and John Cleese

Acted in TV shows

```
In [236]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='United Kingdom') & (df_c['cast']!='nan') & (df_c['type']=='TV Show')]
data_x=data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Actor in United Kingdom for TV Show")
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```



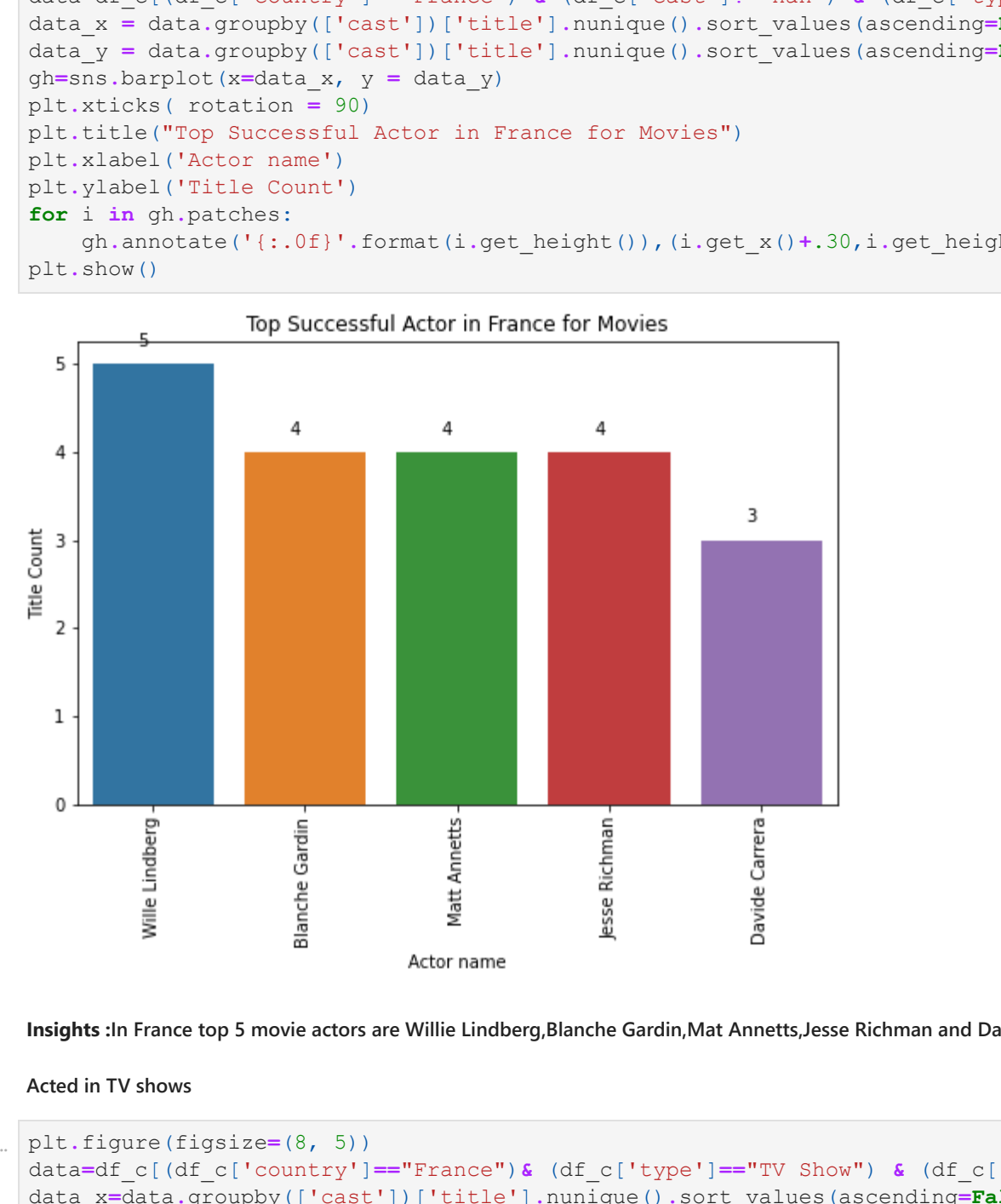
Insights in United Kingdom top 5 TV show actors are David Attenborough,Greg Davies,John Cleese,Hugh Bonneville and Eric Idle

Recommendation in United Kingdom Netflix should produce more movies acted by Terry Jones,Eric Idle,Michael Palin,Terry Gilliam and John Cleese and more TV shows acted by David Attenborough,Greg Davies,John Cleese,Hugh Bonneville and Eric Idle

Top 5 actors in Canada

Acted in movies

```
In [237]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='Canada') & (df_c['cast']!='nan') & (df_c['type']=='Movie')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Actor in Canada for Movies")
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```



Insights in Canada top 5 movie actors are John Paul Tremblay,RobWilliams,John Dunsworth,Mike Smith and Michela Luci

Acted in TV shows

```
In [238]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='Canada') & (df_c['type']=='TV Show') & (df_c['cast']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Actor in Canada for TV Show")
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```



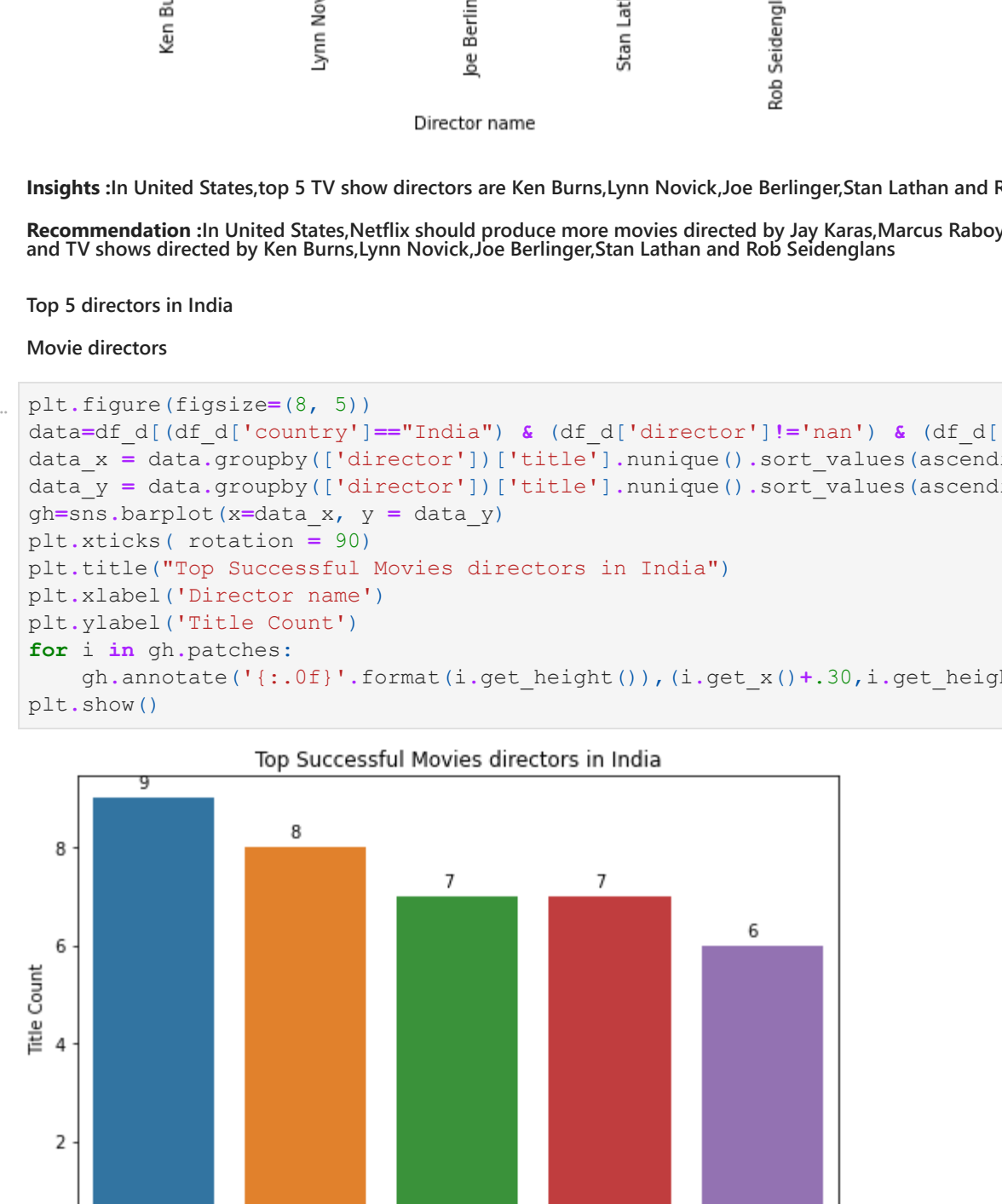
Insights in Canada top 5 TV show actors are John Paul Tremblay,Rob Wells,Mike Smith,Jordan Clark and Trevor Tordjman

Recommendation in Canada Netflix should produce more movies acted by John Paul Tremblay,RobWilliams,John Dunsworth,Mike Smith and Michela Luci and more TV shows acted by John Paul Tremblay,Rob Wells,Mike Smith,Jordan Clark and Trevor Tordjman

Top 5 actors in France

Acted in movies

```
In [239]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='France') & (df_c['cast']!='nan') & (df_c['type']=='Movie')]
data_x = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['cast'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Actor in France for Movies")
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```



Insights in France top 5 movie actors are Willie Lindberg,Blanche Gardin,Mat Annetts,Jesse Richman and Davide Carrera

Acted in TV shows

```
In [240]: plt.figure(figsize=(8, 5))
data=df_c[(df_c['country']=='France') & (df_c['type']=='TV Show') & (df_c['cast']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Actor in France for TV Show")
plt.xlabel('Actor name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```



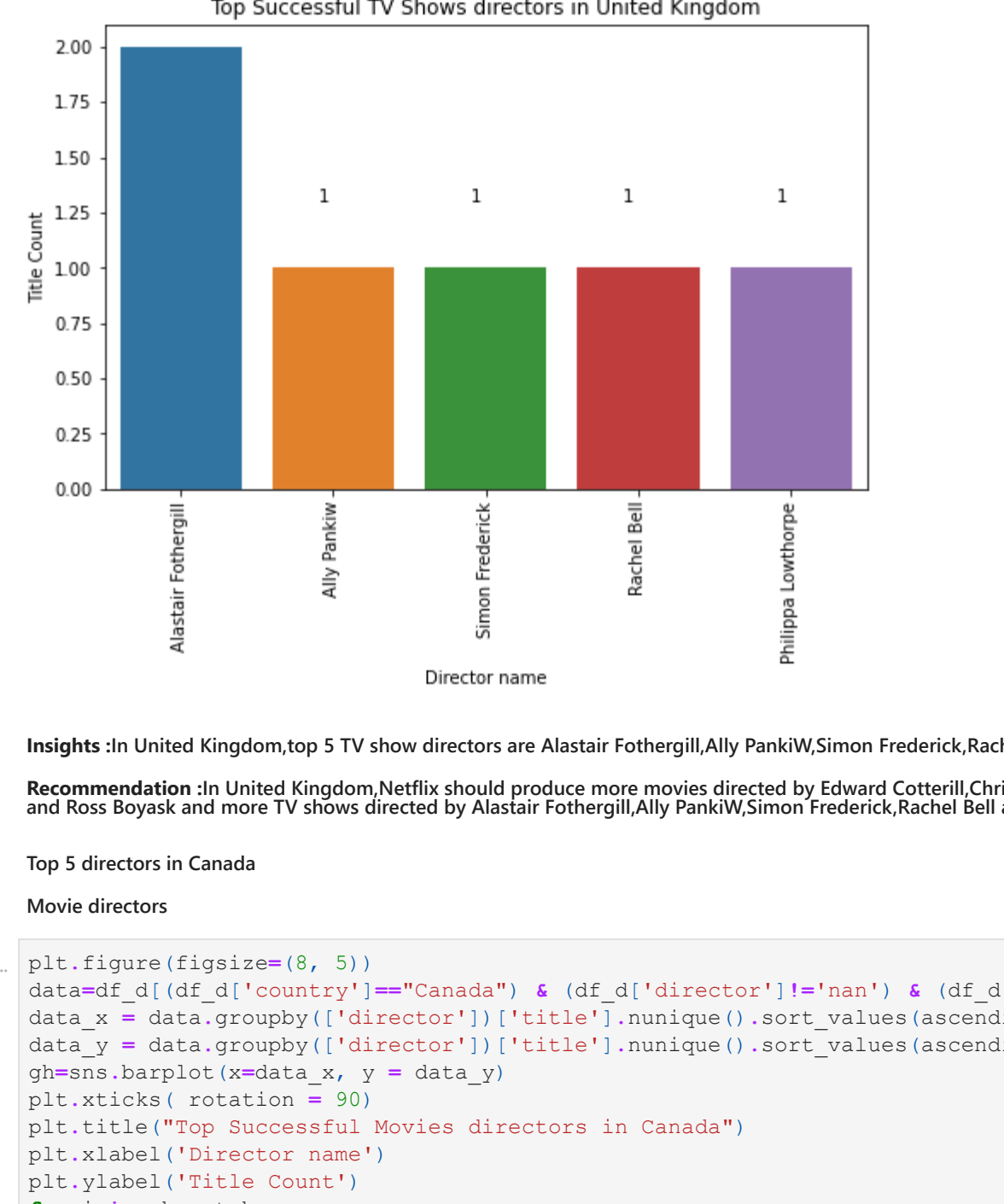
Insights in France top 5 TV show actors are Julia Platon,Kate Moran,Suzanne Clement,Louise Coldefy and Hugo Becker

Recommendation in France Netflix should produce more movies acted by Willie Lindberg,Blanche Gardin,Mat Annetts,Jesse Richman and Davide Carrera and more TV shows acted by Julia Platon,Kate Moran,Suzanne Clement,Louise Coldefy and Hugo Becker

Top 5 directors in United States

Movie directors

```
In [241]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='United States') & (df_d['director']!='nan') & (df_d['type']=='Movie')]
data_x = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Movies directors in United States")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```



Insights in United States top 5 movie directors are Jay Karas,Marcus Raboy,Jay Chapman,Shannon Hartman and Troy Miller

TV shows directors

```
In [242]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='United States') & (df_d['type']=='TV Show') & (df_d['director']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful TV Shows directors in United States")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```



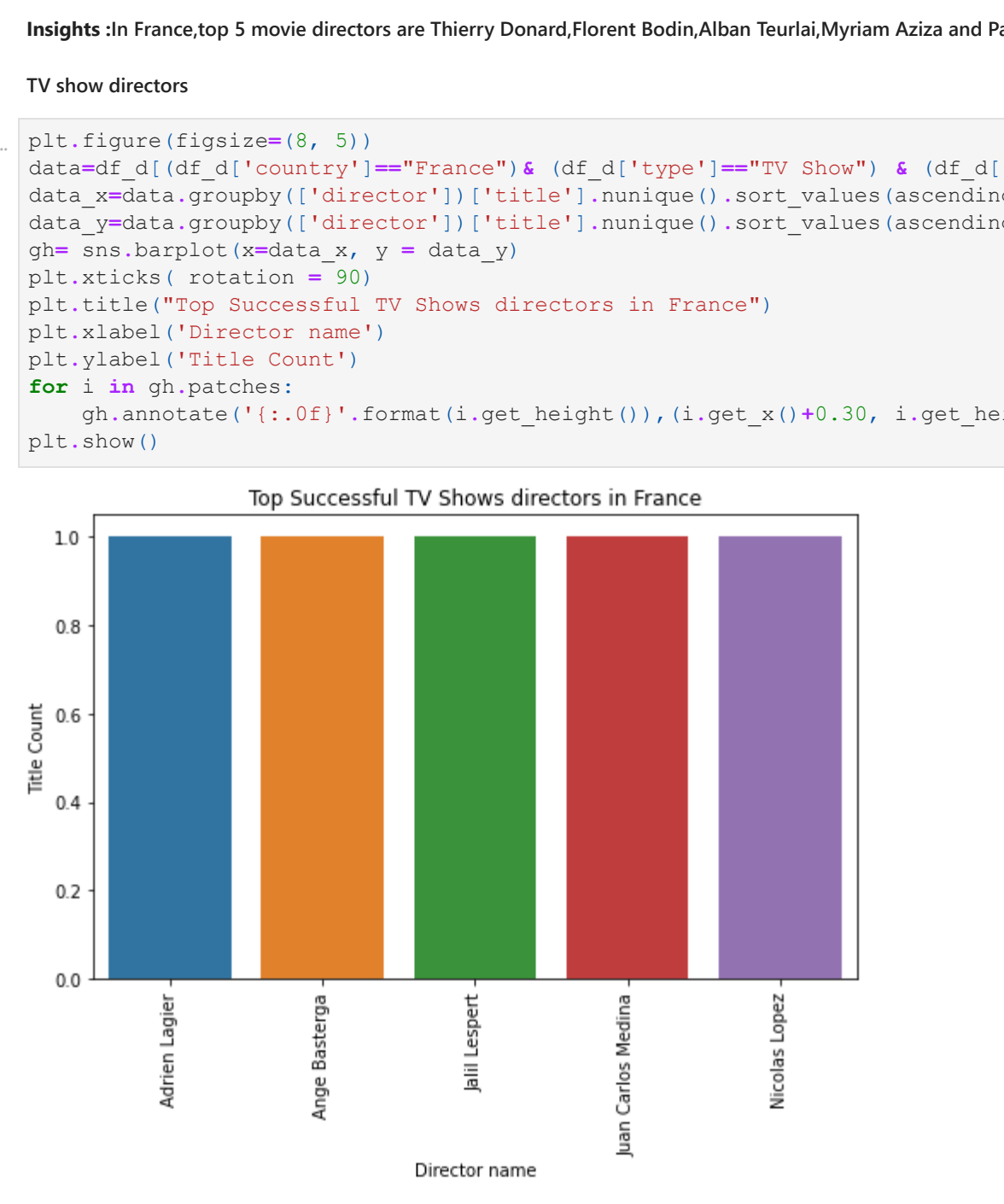
Insights in United States top 5 TV show directors are Ken Burns,Lynn Novick,Joe Berlinger,Stan Lathan and Rob Seidenglanz

Recommendation in United States Netflix should produce more movies directed by Jay Karas,Marcus Raboy,Jay Chapman,Shannon Hartman and Troy Miller and TV shows directed by Ken Burns,Lynn Novick,Joe Berlinger,Stan Lathan and Rob Seidenglanz

Top 5 directors in India

Movie directors

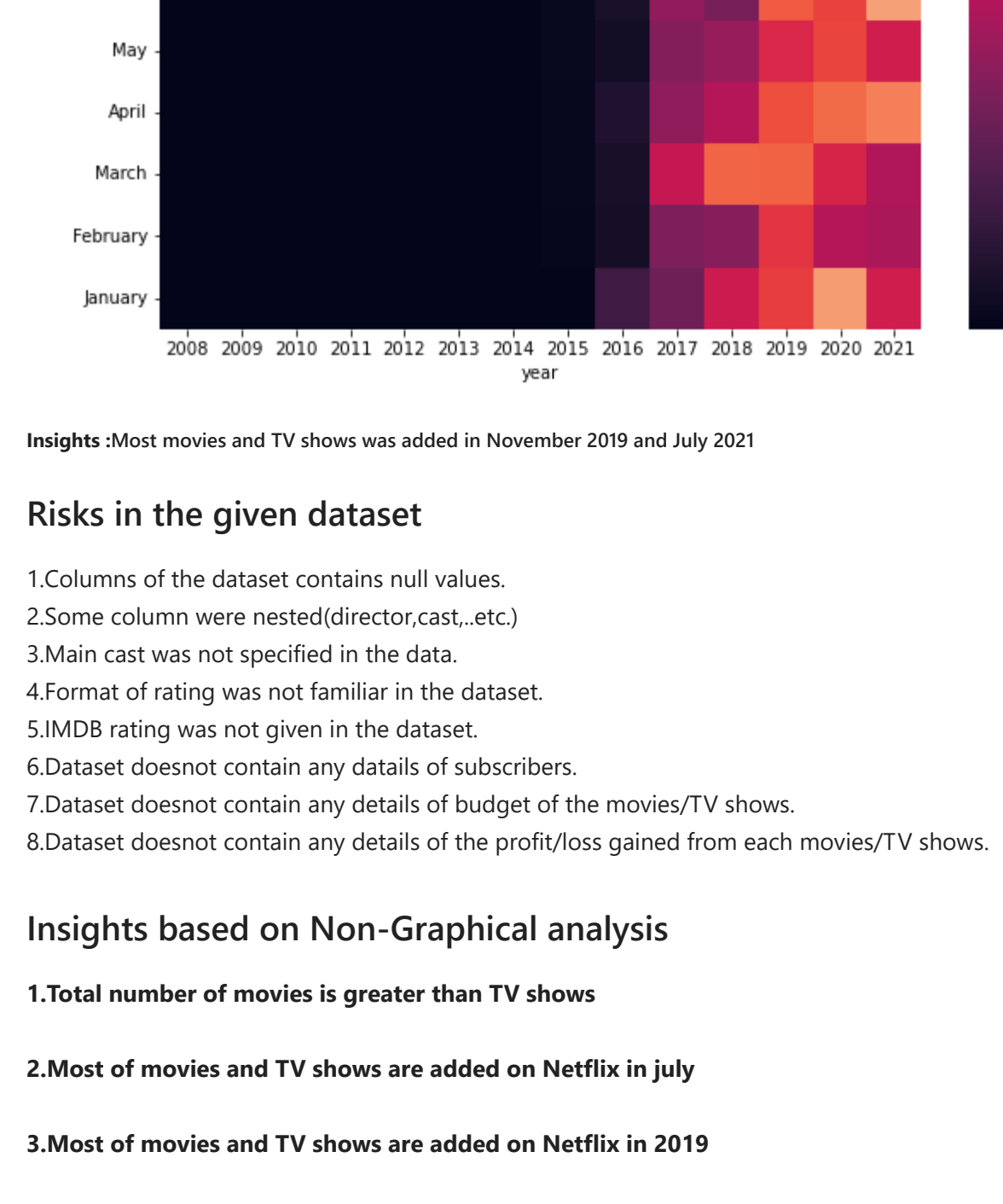
```
In [243]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='India') & (df_d['director']!='nan') & (df_d['type']=='Movie')]
data_x = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Movies directors in India")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```



Insights in India top 5 Movie directors are David Dhawan,Anurag Kashyap,Dibakar Banerjee,Ram Gopal Varma and Sooraj R Barjatya

TV shows directors

```
In [244]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='India') & (df_d['type']=='TV Show') & (df_d['director']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful TV Shows directors in India")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```



Insights in India top 5 TV shows directors are Gautham Vasudev Menon,Arvind Swamy,Rathindran R Prasad,Vignesh Shivan and Vetri Maran

Recommendation in India Netflix should produce more movies directed by David Dhawan,Anurag Kashyap,Dibakar Banerjee,Ram Gopal Varma and Sooraj R Barjatya and more TV Shows directed by Gautham Vasudev Menon,Arvind Swamy,Rathindran R Prasad,Vignesh Shivan and Vetri Maran

Top 5 directors in United Kingdom

Movie directors

```
In [245]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='United Kingdom') & (df_d['director']!='nan') & (df_d['type']=='Movie')]
data_x = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Movies directors in United Kingdom")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```



Insights in United Kingdom top 5 movies directors are Edward Cortelli,Chris Howe,Ian MacNaughton,Michael Winterbottom and Ross Boyask

TV shows directors

```
In [246]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='United Kingdom') & (df_d['type']=='TV Show') & (df_d['director']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful TV Shows directors in United Kingdom")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```


Insights in United Kingdom top 5 TV show directors are Alastair Fothergill,Ally Patten,Simon Frederick,Rachel Bell and Philippa Lowthorpe

Recommendation in United Kingdom Netflix should produce more movies directed by Edward Cortelli,Chris Howe,Ian MacNaughton,Michael Winterbottom and Ross Boyask and more TV Shows directed by Alastair Fothergill,Ally Patten,Simon Frederick,Rachel Bell and Philippa Lowthorpe

Top 5 directors in Canada

Movie directors

```
In [247]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='Canada') & (df_d['director']!='nan') & (df_d['type']=='Movie')]
data_x = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Movies directors in Canada")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```


Insights in Canada top 5 movie directors are Justin G Dyck,Mike Clattenburg,Mike Smith,Barry Avrich and John Paul Tremblay

TV show directors

```
In [248]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='Canada') & (df_d['type']=='TV Show') & (df_d['director']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful TV Shows directors in Canada")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```


Insights in Canada top 5 TV show directors are Gary Howsan,John Paul Tremblay,Mick Grogan,Mike Smith and Norm Hiscock

Recommendation in Canada Netflix should produce more movies directed by Justin G Dyck,Mike Clattenburg,Mike Smith,Barry Avrich and John Paul Tremblay and more TV shows directed by Gary Howsan,John Paul Tremblay,Mick Grogan,Mike Smith and Norm Hiscock

Top 5 directors in France

Movie directors

```
In [249]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='France') & (df_d['director']!='nan') & (df_d['type']=='Movie')]
data_x = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y = data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful Movies directors in France")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.2))
plt.show()
```


Insights in France top 5 movie directors are Thierry Donard,Florent Bodin,Alban Teurla,Myriam Aziza and Pascal Blanchard

TV show directors

```
In [250]: plt.figure(figsize=(8, 5))
data=df_d[(df_d['country']=='France') & (df_d['type']=='TV Show') & (df_d['director']!='nan') ]
data_x=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
data_y=data.groupby(['director'])['title'].nunique().sort_values(ascending=False).head().index
gh=sns.barplot(x=data_x, y = data_y)
plt.xticks(rotation = 90)
plt.title("Top Successful TV Shows directors in France")
plt.xlabel('Director name')
plt.ylabel('Title Count')
for i in gh.patches:
    gh.annotate('{}'.format(i.get_height()),(i.get_x()+0.30,i.get_height()+0.3))
plt.show()
```


Insights in France top 5 TV show directors are Adrien Lagier,Ange Basterga,Jali Laspert,Juan Carlos Medina and Nicolas Lopez

Recommendation in France Netflix should produce more movies directed by Thierry Donard,Florent Bodin,Alban Teurla,Myriam Aziza and Pascal Blanchard and more TV shows directed by Adrien Lagier,Ange Basterga,Jali Laspert,Juan Carlos Medina and Nicolas Lopez

Movie and TV shows updates per months

```
In [251]: date = df[['date_added']].dropna()
date['month'] = date['date_added'].apply(lambda x: x.strftime('%b').split(' ')[0])
date['year'] = date['date_added'].apply(lambda x: x.split('-')[1])
months = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December']
dg = date.groupby('year')['month'].value_counts().unstack().fillna(0).fillna(0)(months)?
plt.figure(figsize=(10, 8))
sns.heatmap(dg)
plt.title('Movie and TV shows updates per months')
plt.show()
```


Insights :Most movies and TV shows was added in November and July 2021

Risks in the given dataset

- 1.Counts of the dataset contains null values.
- 2.Some column were nested (director.cast,etc)
- 3.Main cast was not specified in the data.
- 4.Format of rating was not given in the dataset.
- 5.IMDB rating was not given in the dataset.
- 6.Dataset doesn't contain any datails of subscribers.
- 7.Dataset doesn't contain any details of budget of the movies/TV shows.
- 8.Dataset doesn't contain any datails of the profit/loss gained from each movies/TV shows.

Insights based on Non-Graphical analysis

- 1.Total number of movies is greater than TV shows
- 2.Most of movies and TV shows are addet on Netflix in July
- 3.Most of movies and TV shows are added on Netflix in 2019
- 4.Most number of movies and TV shows are released in 2018
- 5.Most number of movies and TV shows are rated as TV-MA

Insights based on visual analysis

- 1.Total number of movies is more than TV shows
 - 2.Till 2018 the trend of movies was higher than that of TV shows.But after 2018 movies is showing decrease in trend.In 2021 TV shows trend is higher than that of movies.
 - 3.Most of Movies/TV shows are released between the year 2018 and 2020
 - 4.Most number of movies and TV shows are released in 2018
 - 5.Most number of movies and TV shows on Netflix are rated as TV-MA
 - 6.Most of movies have 90 minutes duration
 - 7.Most number of TV shows have 1 season duration.
 - 8.Most number of movies and TV shows are directed by Rajiv Chilaka
 - 9.Most number of movies and TV shows are acted by Anupam Kher
 - 10.Most number of movies and TV shows are released in United States
 - 11.International Movies are the most among listed_in
 - 12.Dramas type of genre is most watched in United States
 - 13.International movies type of genre is most watched in India
 - 14.British TV Shows type of genre is most watched in United Kingdom
 - 15.Comedy type of genre is most watched in Canada
 - 16.International movies type of genre is most watched in France.
 - 17.In United States top 5 movie actors are Adam Sandler,Samual L.Jackson,Laura Bailey,Molly Shannon and Nicolas Cage
 - 18.In United States top 5 TV shows actors are Mike Colter,Grey Griffin,Kevin Michael,Steven Yeun and Kari Wahlgren
 - 19.In India top 5 movie actors are Anupam Kher,Shah Rukh Khan,Naseeruddin Shan,Amitabh Bachchan and Akshay Kumar
 - 20.In India top 5 TV show actors are Nishka Raheja,Rajesh Kava,Prakash Raj,Mona Ambegaonkar and Vineet Kumar
 - 21.In United Kingdom top 5 movie actors are Terry Jones,Eric Idle,Michael Palin,Terry Gillian and John Cleese
 - 22.In United Kingdom top 5 TV show actors are David Attenborough,Greg Davies,John Cleese,Hugh Bonneville and Eric Idle
 - 23.In Canada top 5 movie actors are John Paul Tremblay,Robwells,John Dunsworth,Mike Smith and Michela Luci
 - 24.In Canada top 5 TV show actors are John Paul Tremblay,Rob Wells,Mike Smith,Jordan Clark and Trevor Tordjman
 - 25.In France top 5 movie actors are Willie Lindberg,Blanche Gardin,Mat Annetts,Jesse Richman and Davide Carrera
 - 26.In France top 5 TV show actors are Julia Piaton,Kate Moran,Suzanne Clement,Louise Coldefy and Hugo Becker
 - 27.In United States,top 5 movie directors are Jay Karas,Marcus Raboy,Jay Chapman,Shannon Hartman and Troy Miller
 - 28.In United States,top 5 TV show directors are Ken Burns,Lynn Novick,Joe Berlinger,Stan Lathan and Rob Seidenglans
 - 29.In India,top 5 Movie directors are David Dhavan,Anurag Kashyap,Dibakar Banerjee,Ram Gopal Varma and Sooraj R.Barjatya
 - 30.In India,top 5 TV shows directors are Gautham Vasudev Menon,Arvind Swamy,Rathindran R Prasad,Vignesh Shivan and Vetri Maaran
 - 31.In United Kingdom,top 5 movies directors are Edward Cotterill,Chris Howe,Ian MacNaughton,Michael Winterbottom and Ross Boyask
 - 32.In United Kingdom,top 5 TV show directors are Alastair Fothergill,Ally PankiW,Simon Frederick,Rachel Bell and Philippa Lowthorpe
 - 33.In Canada,top 5 movie directors are Justin G.Dyck,Mike Clattenburg,Mike Smith,Barry Avrich and John Paul Tremblay
 - 34.In Canada,top 5 TV show directors are Gray Howsam,John Paul Tremblay,Mick Grogan,Mike Smith and Norm Hiscock
 - 35.In France,top 5 movie directors are Thierry Donard,Florent Bodin,Alban Teurlai,Myriam Aziza and Pascal Blanchard
 - 36.In France,top 5 TV show directors are Adrien Lagier,Ange Basterga,Jalil Lespert,Juan Carlos Medina and Nicolas Lopez
 - 37.Most movies and TV shows was added in November 2019 and July 2021
- ### Recommendations
- 1.Netflix should produce more TV shows than movies as per the trend.
 - 2.Netflix should produce more TV-MA rated movies and TV shows.
 - 3.Netflix should produce movies having 90 minutes duration.
 - 4.Netflix should produce TV shows having 1 season duration.
 - 5.Netflix should produce more Dramas in United States.
 - 6.Netflix should produce more International movies in India.
 - 7.Netflix should produce more British TV Shows in United Kingdom.
 - 8.Netflix should produce more Comedy type of genre in Canada.
 - 9.Netflix should produce more International movies in France.
 - 10.In United States,Netflix should produce more movies acted by Adam Sandler,Samual L.Jackson,Laura Bailey,Molly Shannon and Nicolas Cage and more TV shows acted by Mike Colter,Grey Griffin,Kevin Michael,Steven Yeun and Kari Wahlgren.
 - 11.In India,Netflix should produce more movies acted by Anupam Kher,Shah Rukh Khan,Naseeruddin Shan,Amitabh Bachchan and Akshay Kumar and more TV shows acted by Nishka Raheja,Rajesh Kava,Prakash Raj,Mona Ambegaonkar and Vineet Kumar.
 - 12.In United Kingdom,Netflix should produce more movies acted by Terry Jones,Eric Idle,Michael Palin,Terry Gillian and John Cleese and more TV shows acted by David Attenborough,Greg Davies,John Cleese,Hugh Bonneville and Eric Idle.
 - 13.In Canada,Netflix should produce more movies acted by John Paul Tremblay,Robwells,John Dunsworth,Mike Smith and Michela Luci and more TV shows acted by John Paul Tremblay,Rob Wells,Mike Smith,Jordan Clark and Trevor Tordjman.
 - 14.In France,Netflix should produce more movies acted by Willie Lindberg,Blanche Gardin,Mat Annetts,Jesse Richman and Davide Carrera and more TV shows acted by Julia Piaton,Kate Moran,Suzanne Clement,Louise Coldefy and Hugo Becker.
 - 15.In United States,Netflix should produce more movies directed by Jay Karas,Marcus Raboy,Jay Chapman,Shannon Hartman and Troy Miller and TV shows directed by Ken Burns,Lynn Novick,Joe Berlinger,Stan Lathan and Rob Seidenglans.
 - 16.In India,Netflix should produce more movies directed by David Dhavan,Anurag Kashyap,Dibakar Banerjee,Ram Gopal Varma and Sooraj R.Barjatya and more TV shows directed by Gautham Vasudev Menon,Arvind Swamy,Rathindran R Prasad,Vignesh Shivan and Vetri Maaran.
 - 17.In United Kingdom,Netflix should produce more movies directed by Edward Cotterill,Chris Howe,Ian MacNaughton,Michael Winterbottom and Ross Boyask and more TV shows directed by Alastair Fothergill,Ally PankiW,Simon Frederick,Rachel Bell and Philippa Lowthorpe.
 - 18.In Canada,Netflix should produce more movies directed by Justin G.Dyck,Mike Clattenburg,Mike Smith,Barry Avrich and John Paul Tremblay and more TV shows directed by Gray Howsam,John Paul Tremblay,Mick Grogan,Mike Smith and Norm Hiscock.
 - 19.In France,Netflix should produce more movies directed by Thierry Donard,Florent Bodin,Alban Teurlai,Myriam Aziza and Pascal Blanchard and more TV shows directed by Adrien Lagier,Ange Basterga,Jalil Lespert,Juan Carlos Medina and Nicolas Lopez.
 - 20.United States,India,United Kingdom,Canada,France are the top 5 countries where Netflix is mostly used.Netflix should focus on these countries for getting more profit.