NETFLIX DATA ANALYSIS

PRESENTED BY

Sanket

Aayush Choudhary

Rajbhuwan

Ankit

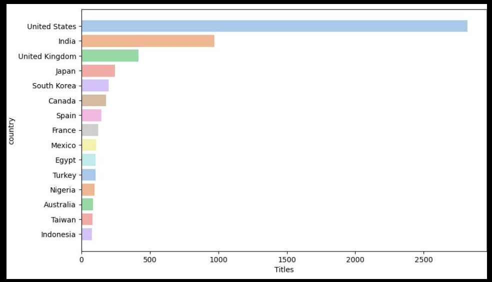
Rohit

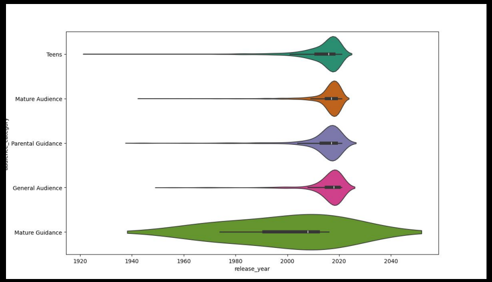
The best Group

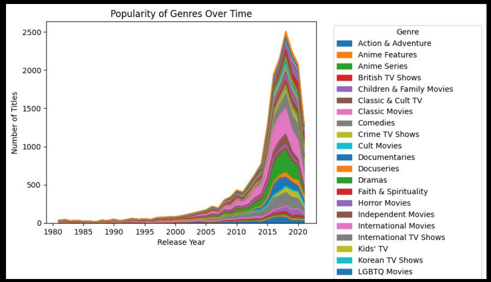
INTRODUCTION

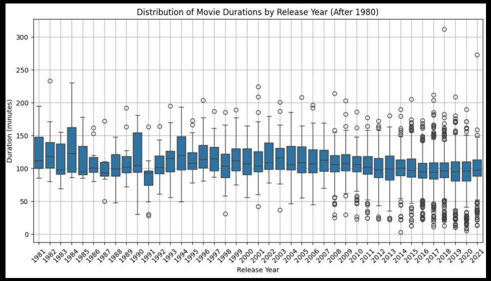
97% for you **20+** 2024

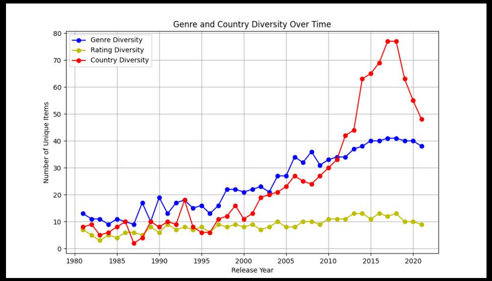
This project aims to analyze various aspects of Netflix's extensive library, including the growth of movies and TV shows, genre distribution, country of origin, content duration, rating categories, release patterns, director and cast involvement, genre popularity over time, and much more.

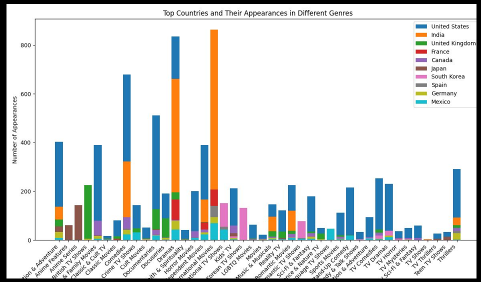


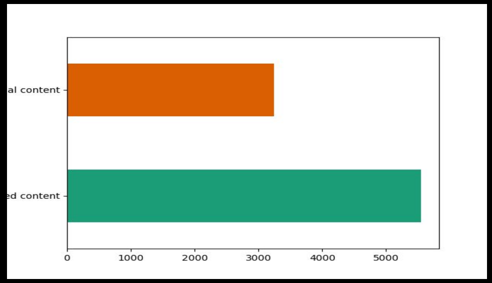


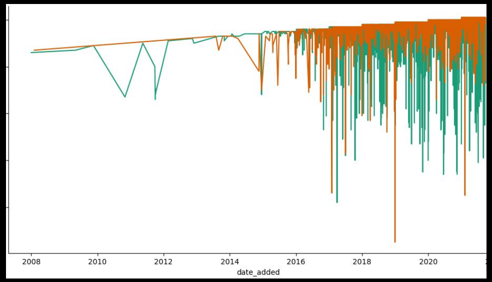


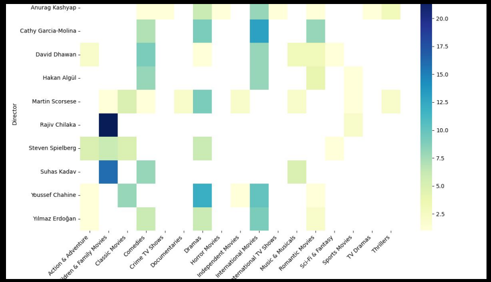












▶ Skip

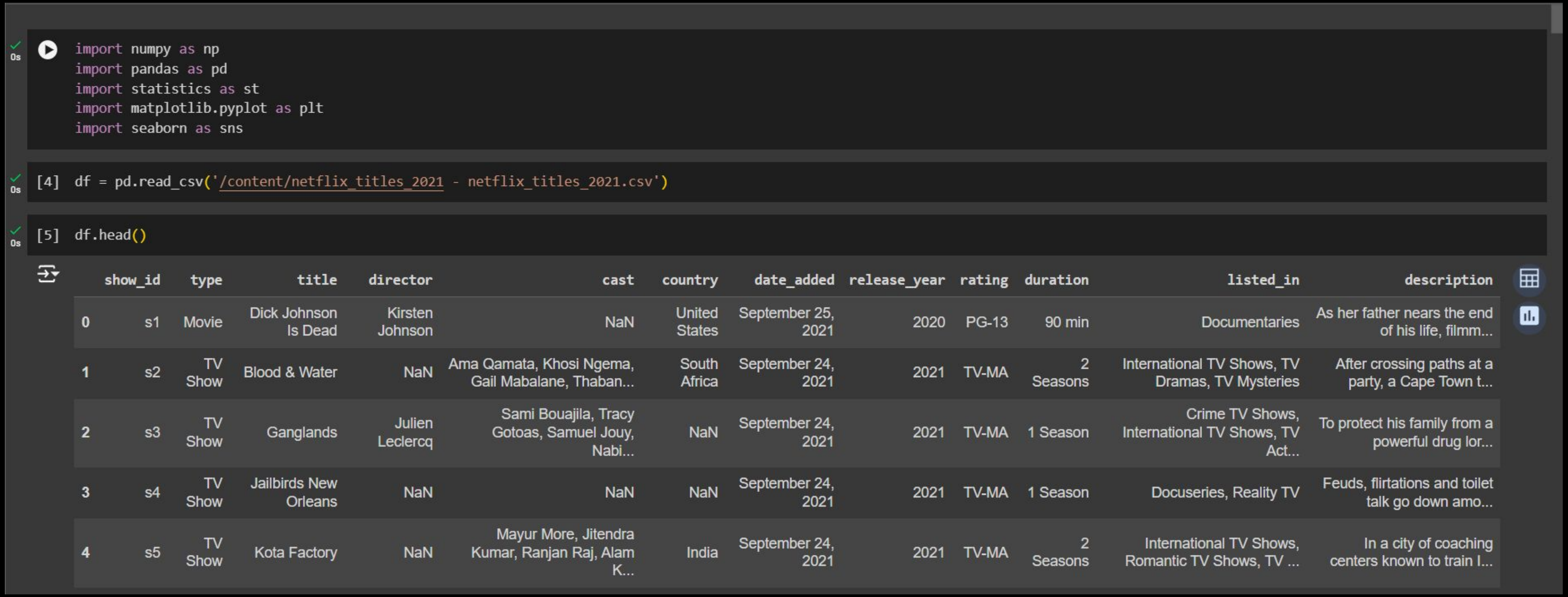
LOADING DATA ....

Libraries and Dataset Used for this project

97% for you 20+ 2024

List of Libraries

1. Pandas
2. NumPy
3. Matplotlib
4. Seaborn
5. Statistics
6. Itertools



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

```
[4] df = pd.read_csv('/content/netflix_titles_2021 - netflix_titles_2021.csv')
```

```
[5] df.head()
```

	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 l...

DATA UNDERSTANDING

- Number of rows and columns: [8807,9]

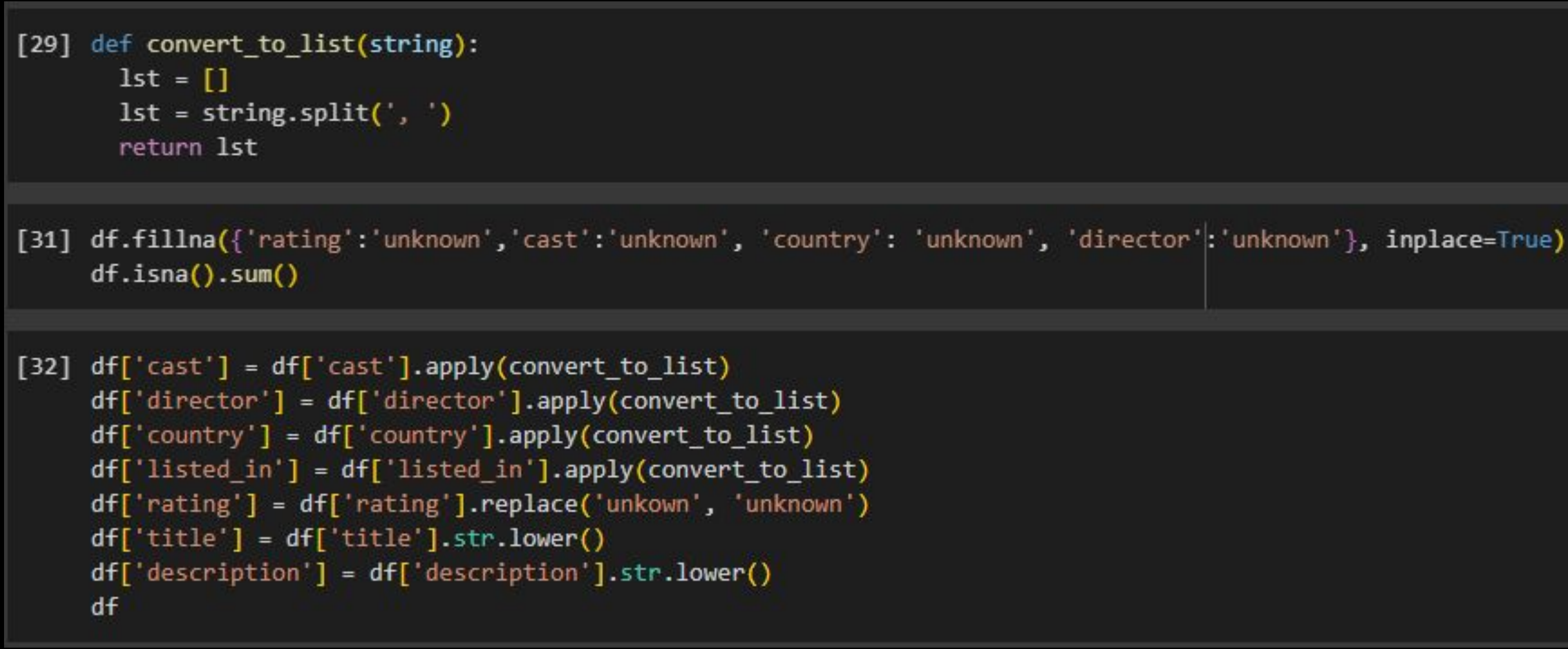
  Data types :

  - **show_id** : object *A unique identifier for each title.*
  - **type** : object *The category of the title, which is either 'Movie' or 'TV Show'.*
  - **title** : object *The name of the movie or TV show.*
  - **director** : object The director(s) of the movie or TV show. (Contains null values for some entries, especially TV shows where this information might not be applicable.)
  - **cast**: object *The list of main actors/actresses in the title. (Some entries might not have this information.)*
  - **country**: object *The country or countries where the movie or TV show was produced.*
  - **date_added**: datetime64[ns] *The date the title was added to Netflix.*
  - **release_year**: int64 *The year the movie or TV show was originally released.*
  - **rating**: object *The age rating of the title.*
  - **duration**: object *The duration of the title, in minutes for movies and seasons for TV shows.*
  - **listed_in**: object *The genres the title falls under.*
  - **description**: object *A brief summary of the title.*

International Movies
Dramas International
Movies Dramas
Movies Comedies
Romantic Movies
Movies Documentaries
Movies Action
Family Movies
Dramas Independent
Comedies Dramas
Documentaries International
Action Adventure
Children Family
Movies Romantic
Movies International
Independent Movies
Movies Thrillers
Movies Children
Music Musicals
Comedies International
Horror Movies
Adventure Dramas
Adventure Comedies
Sci Fi
Fi Fantasy
Sports Movies
Movies Music
Movies Horror
Movies Stand
Comedies Comedies
LGBTQ Movies
Fantasy Dramas
Fantasy Action
Cult Movies
Comedies Horror
Thrillers Dramas
Thrillers Documentaries
Documentaries Music
Documentaries Sports
Comedy Documentaries
Faith Spirituality
Anime Features
Dramas Comedies
Thrillers Action
Thrillers Children
Comedy Dramas
Classic Movies
Movies Independent
Movies Sci
Documentaries Documentaries
Musicals Dramas
Comedies Romantic
Thrillers Comedies
Musicals Action
Musicals Documentaries
Comedy
Stand

▶ Main Slide

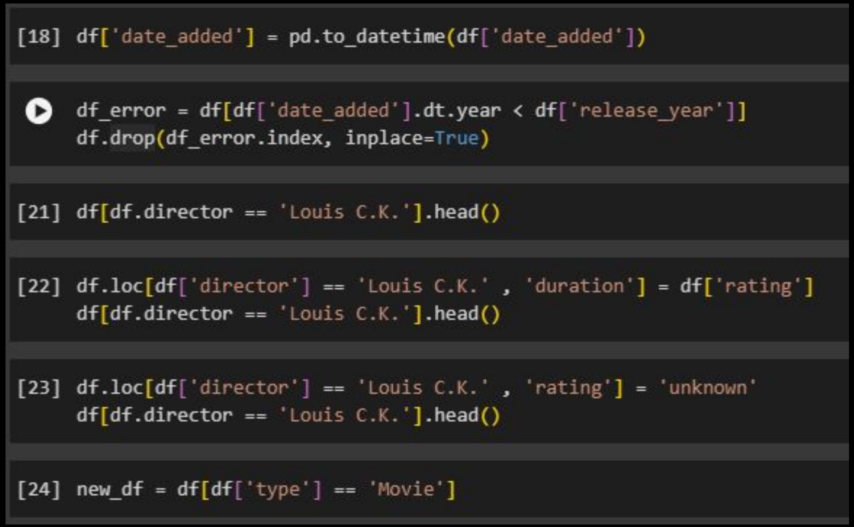
DATA CLEANING



```
[29] def convert_to_list(string):
       lst = []
       lst = string.split(', ')
       return lst
```

```
[31] df.fillna({'rating':'unknown','cast':'unknown', 'country': 'unknown', 'director':'unknown'}, inplace=True)
     df.isna().sum()
```

```
[32] df['cast'] = df['cast'].apply(convert_to_list)
     df['director'] = df['director'].apply(convert_to_list)
     df['country'] = df['country'].apply(convert_to_list)
     df['listed_in'] = df['listed_in'].apply(convert_to_list)
     df['rating'] = df['rating'].replace('unkown', 'unknown')
     df['title'] = df['title'].str.lower()
     df['description'] = df['description'].str.lower()
     df
```



```
[18] df['date_added'] = pd.to_datetime(df['date_added'])
```

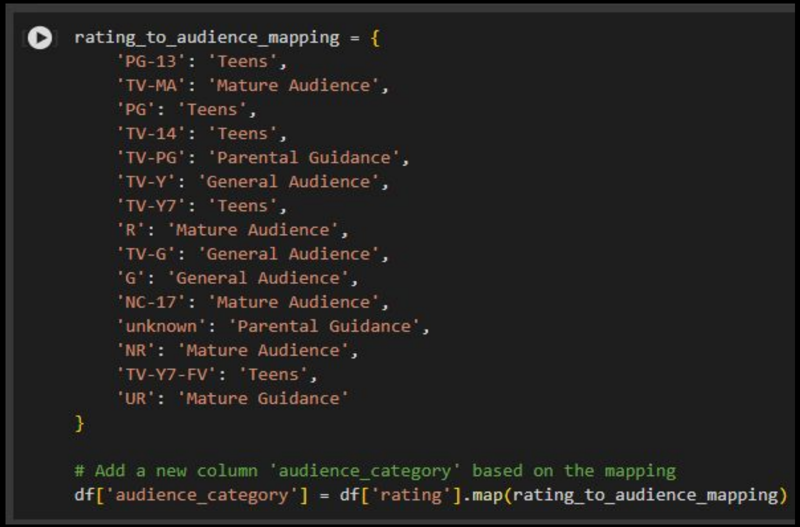
```
     df_error = df[df['date_added'].dt.year < df['release_year']]
     df.drop(df_error.index, inplace=True)
```

```
[21] df[df.director == 'Louis C.K.'].head()
```

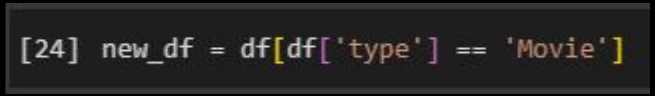
```
[22] df.loc[df['director'] == 'Louis C.K.' , 'duration'] = df['rating']
     df[df.director == 'Louis C.K.'].head()
```

```
[23] df.loc[df['director'] == 'Louis C.K.' , 'rating'] = 'unknown'
     df[df.director == 'Louis C.K.'].head()
```

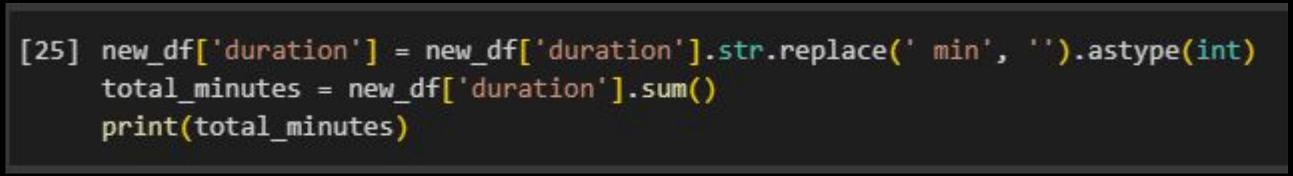
```
[24] new_df = df[df['type'] == 'Movie']
```



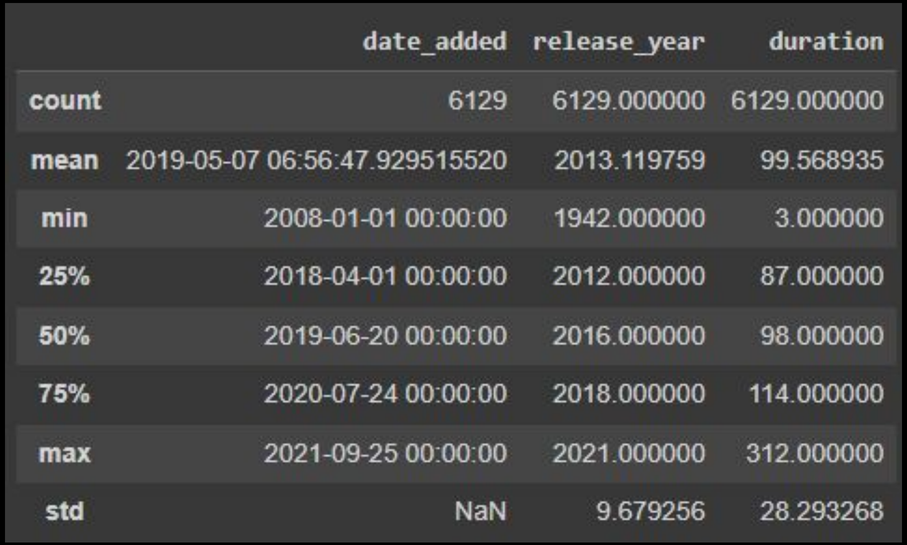
```
rating_to_audience_mapping = {
    'PG-13': 'Teens',
    'TV-MA': 'Mature Audience',
    'PG': 'Teens',
    'TV-14': 'Teens',
    'TV-PG': 'Parental Guidance',
    'TV-Y': 'General Audience',
    'TV-Y7': 'Teens',
    'R': 'Mature Audience',
    'TV-G': 'General Audience',
    'G': 'General Audience',
    'NC-17': 'Mature Audience',
    'unknown': 'Parental Guidance',
    'NR': 'Mature Audience',
    'TV-Y7-FV': 'Teens',
    'UR': 'Mature Guidance'
}

# Add a new column 'audience_category' based on the mapping
df['audience_category'] = df['rating'].map(rating_to_audience_mapping)
```



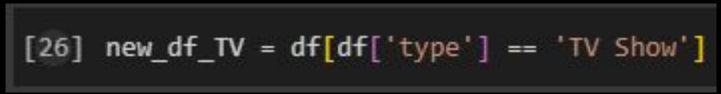
```
[24] new_df = df[df['type'] == 'Movie']
```



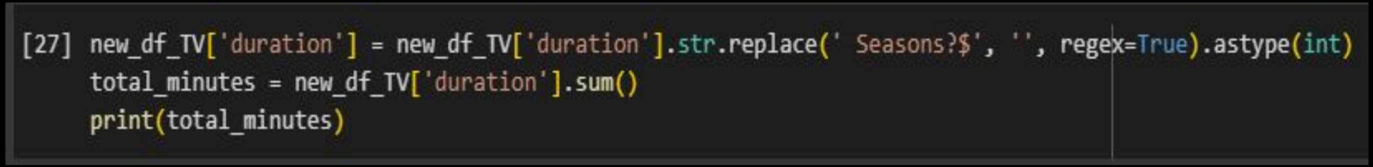
```
[25] new_df['duration'] = new_df['duration'].str.replace(' min', '').astype(int)
     total_minutes = new_df['duration'].sum()
     print(total_minutes)
```



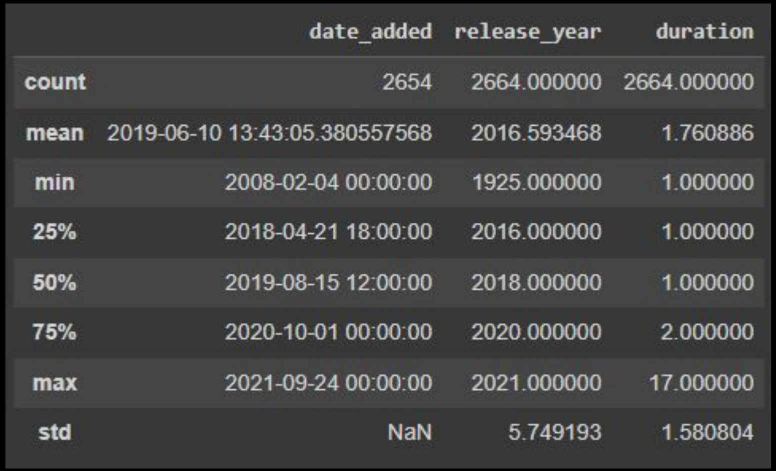
	date_added	release_year	duration
count	6129	6129.000000	6129.000000
mean	2019-05-07 06:56:47.929515520	2013.119759	99.568935
min	2008-01-01 00:00:00	1942.000000	3.000000
25%	2018-04-01 00:00:00	2012.000000	87.000000
50%	2019-06-20 00:00:00	2016.000000	98.000000
75%	2020-07-24 00:00:00	2018.000000	114.000000
max	2021-09-25 00:00:00	2021.000000	312.000000
std	NaN	9.679256	28.293268



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

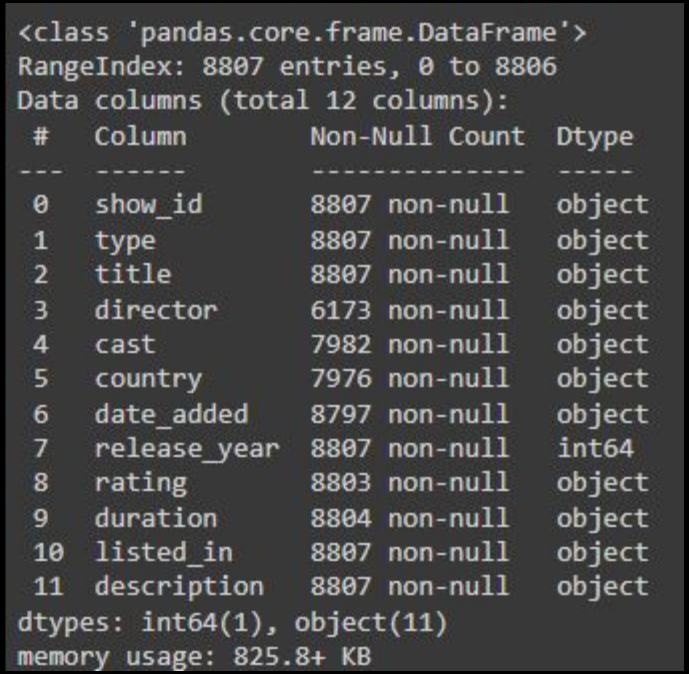


```
[27] new_df_TV['duration'] = new_df_TV['duration'].str.replace(' Seasons?$', '', regex=True).astype(int)
     total_minutes = new_df_TV['duration'].sum()
     print(total_minutes)
```



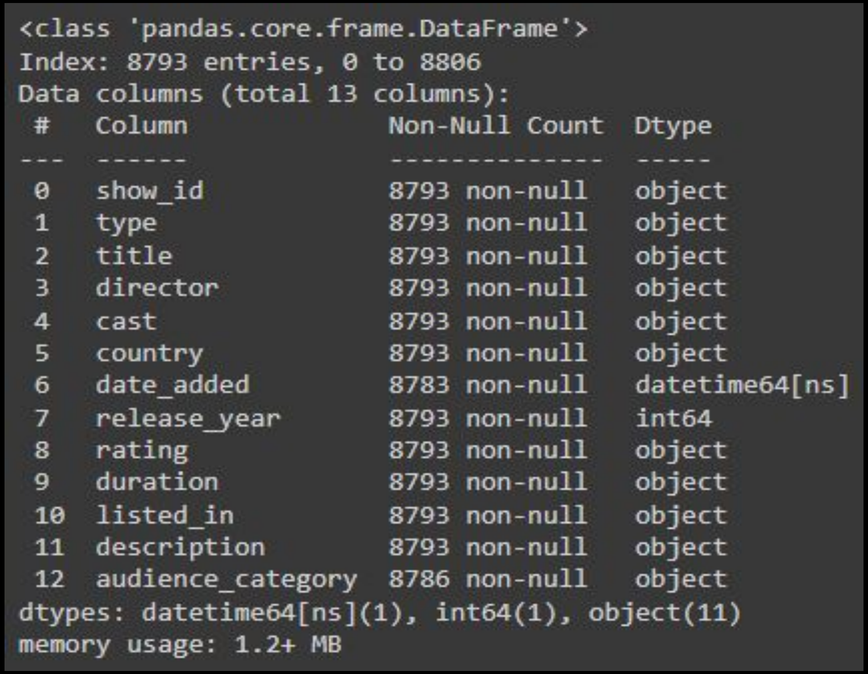
	date_added	release_year	duration
count	2654	2664.000000	2664.000000
mean	2019-06-10 13:43:05.380557568	2016.593468	1.760886
min	2008-02-04 00:00:00	1925.000000	1.000000
25%	2018-04-21 18:00:00	2016.000000	1.000000
50%	2019-08-15 12:00:00	2018.000000	1.000000
75%	2020-10-01 00:00:00	2020.000000	2.000000
max	2021-09-24 00:00:00	2021.000000	17.000000
std	NaN	5.749193	1.580804

BEFORE

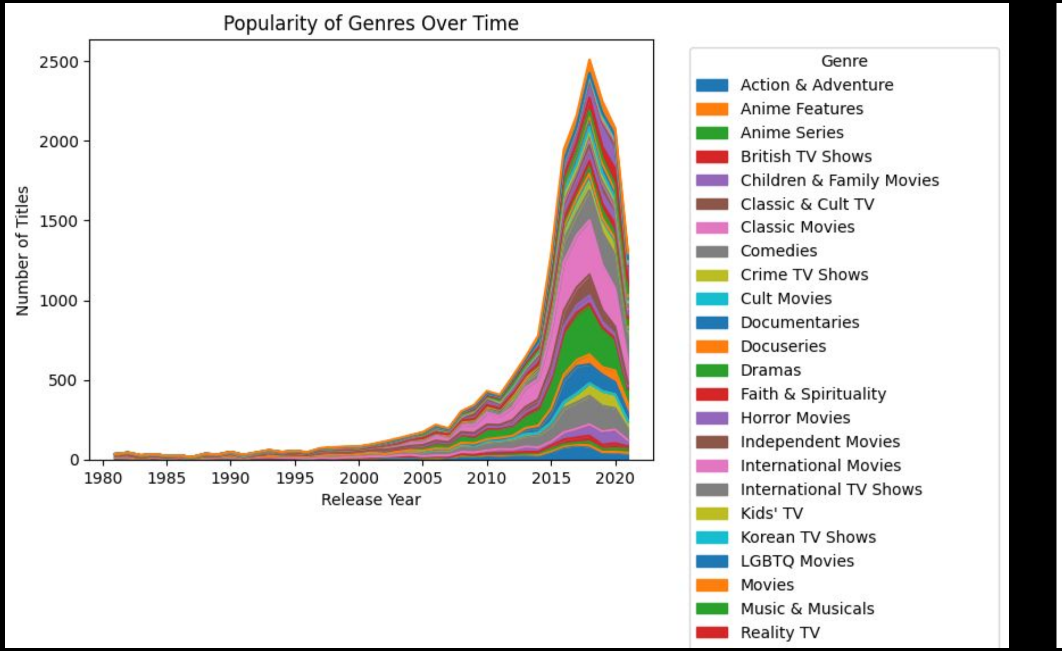


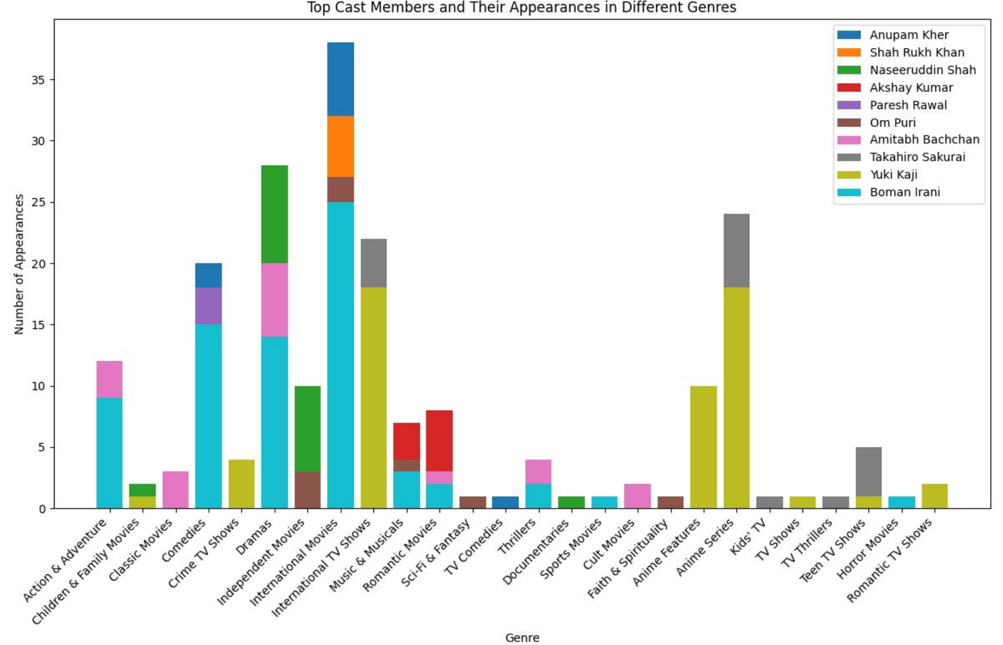
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 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
```

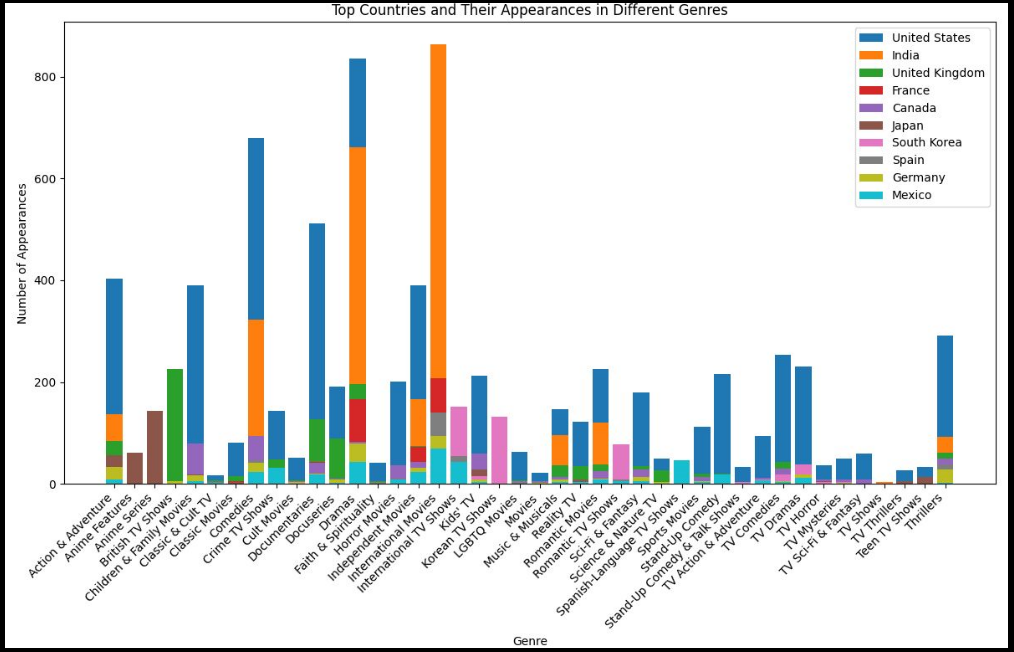
AFTER

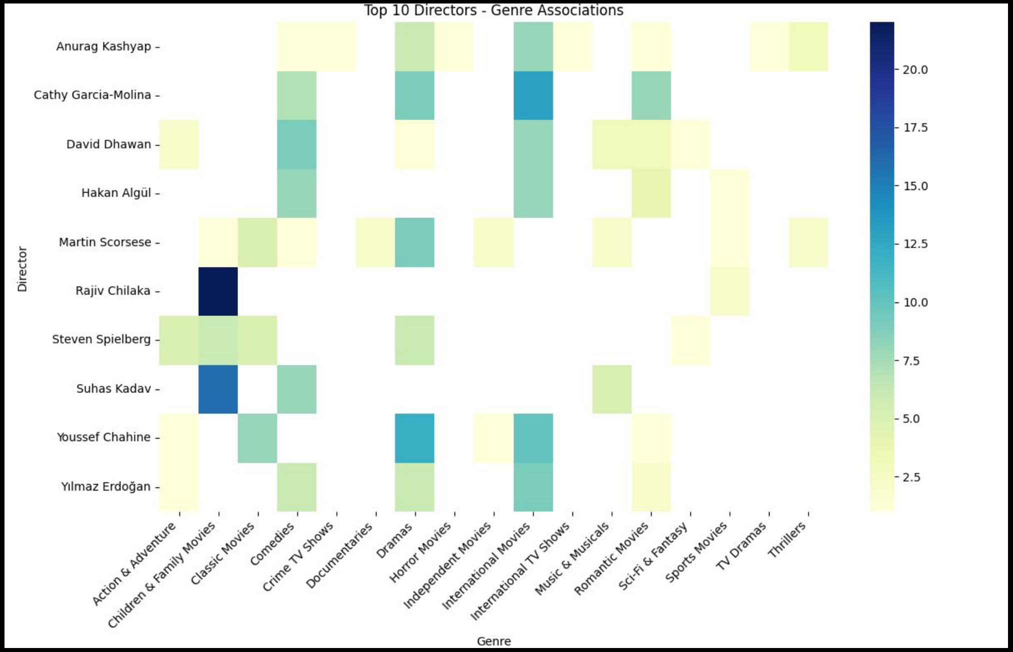


```
<class 'pandas.core.frame.DataFrame'>
Index: 8793 entries, 0 to 8806
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype         
---  ------             --------------  -----         
 0   show_id            8793 non-null   object        
 1   type               8793 non-null   object        
 2   title              8793 non-null   object        
 3   director           8793 non-null   object        
 4   cast               8793 non-null   object        
 5   country            8793 non-null   object        
 6   date_added         8783 non-null   datetime64[ns]
 7   release_year       8793 non-null   int64         
 8   rating             8793 non-null   object        
 9   duration           8793 non-null   object        
 10  listed_in          8793 non-null   object        
 11  description        8793 non-null   object        
 12  audience_category  8786 non-null   object        
dtypes: datetime64[ns](1), int64(1), object(11)
memory usage: 1.2+ MB
```

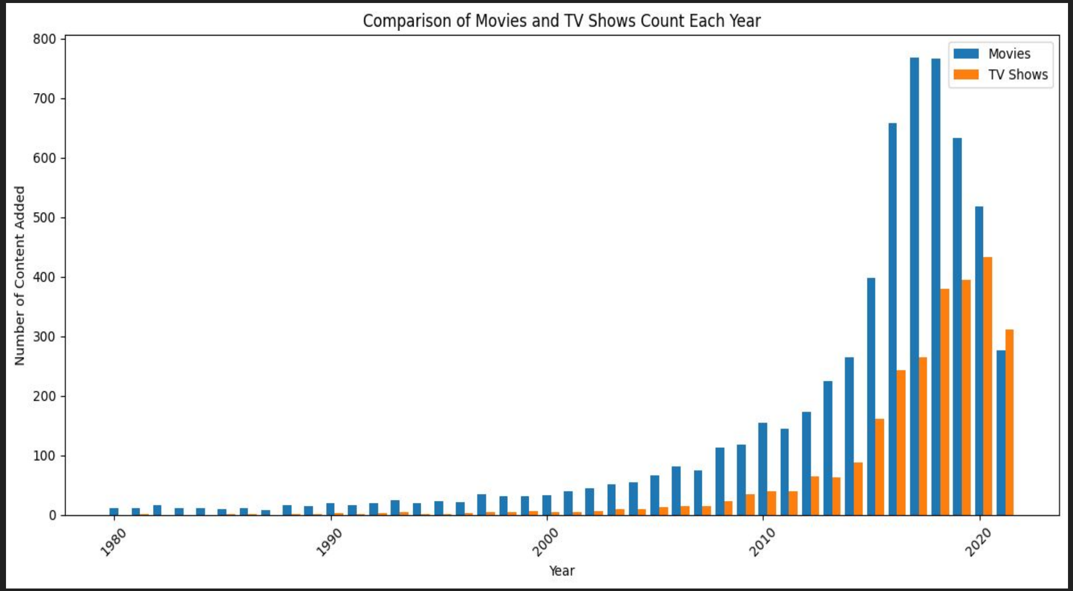

Popularity of Genres Over Time
Number of Titles
2500
2000
1500
1000
500
0
1980
1985
1990
1995
2000
2005
2010
2015
2020
Release Year
Genre
Action & Adventure
Anime Features
Anime Series
British TV Shows
Children & Family Movies
Classic & Cult TV
Classic Movies
Comedies
Crime TV Shows
Cult Movies
Documentaries
Docuseries
Dramas
Faith & Spirituality
Horror Movies
Independent Movies
International Movies
International TV Shows
Kids' TV
Korean TV Shows
LGBTQ Movies
Movies
Music & Musicals
Reality TV

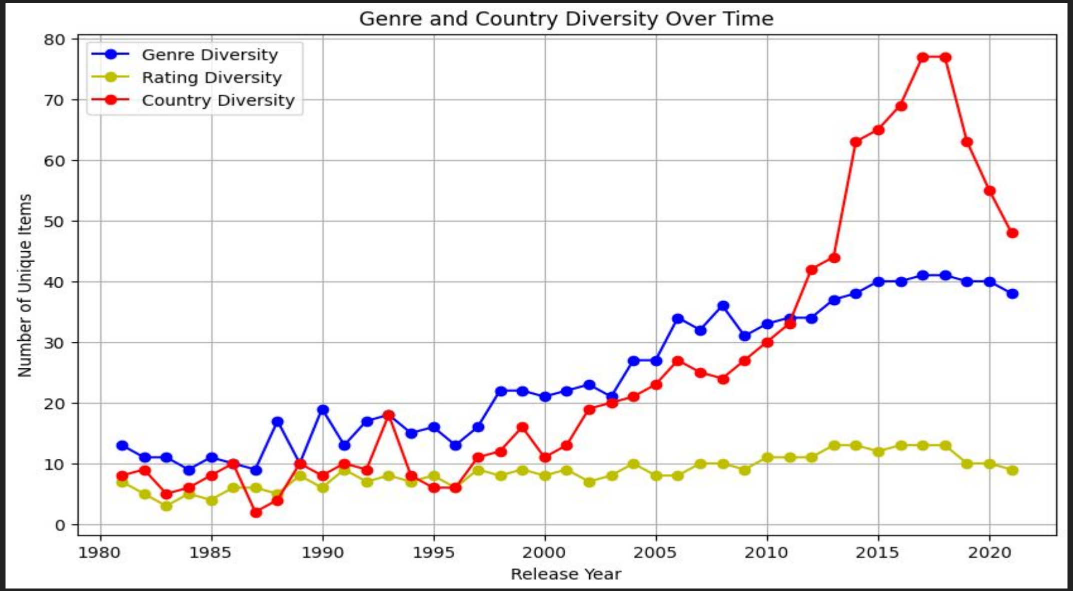

Top Cast Members and Their Appearances in Different Genres
Number of Appearances
35
30
25
20
15
10
5
0
Anupam Kher
Shah Rukh Khan
Naseeruddin Shah
Akshay Kumar
Paresh Rawal
Om Puri
Amitabh Bachchan
Takahiro Sakurai
Yuki Kaji
Boman Irani
Action & Adventure
Children & Family Movies
Classic Movies
Comedies
Crime TV Shows
Dramas
Independent Movies
International Movies
International TV Shows
Music & Musicals
Romantic Movies
Sci-Fi & Fantasy
TV Comedies
Thrillers
Documentaries
Sports Movies
Cult Movies
Faith & Spirituality
Anime Features
Anime Series
Kids' TV
TV Shows
TV Thrillers
Teen TV Shows
Horror Movies
Romantic TV Shows
Genre

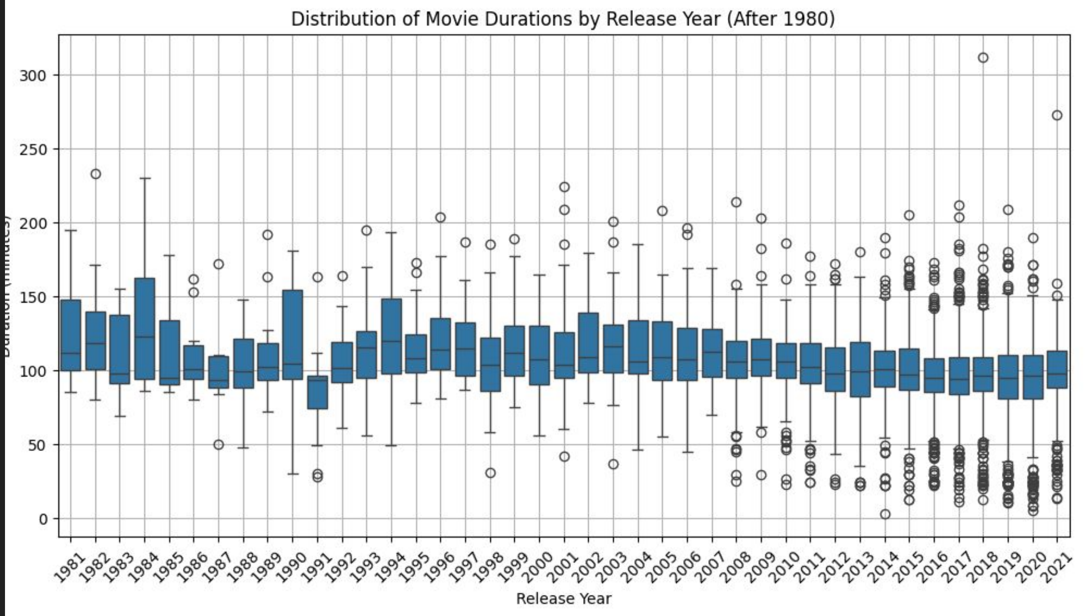

Top Countries and Their Appearances in Different Genres
Number of Appearances
800
600
400
200
0
United States
India
United Kingdom
France
Canada
Japan
South Korea
Spain
Germany
Mexico
Action & Adventure
Anime Features
Anime Series
British TV Shows
Children & Family Movies
Classic & Cult TV
Classic Movies
Comedies
Crime TV Shows
Cult Movies
Documentaries
Docuseries
Dramas
Faith & Spirituality
Horror Movies
Independent Movies
International Movies
International TV Shows
Kids' TV
Korean TV Shows
LGBTQ Movies
Movies
Music & Musicals
Reality TV
Romantic Movies
Romantic TV Shows
Sci-Fi & Fantasy
Science & Nature TV
Spanish-Language TV Shows
Sports Movies
Stand-Up Comedy
Stand-Up Comedy & Talk Shows
TV Action & Adventure
TV Comedies
TV Dramas
TV Horror
TV Mysteries
TV Sci-Fi & Fantasy
TV Shows
TV Thrillers
Teen TV Shows
Thrillers
Genre

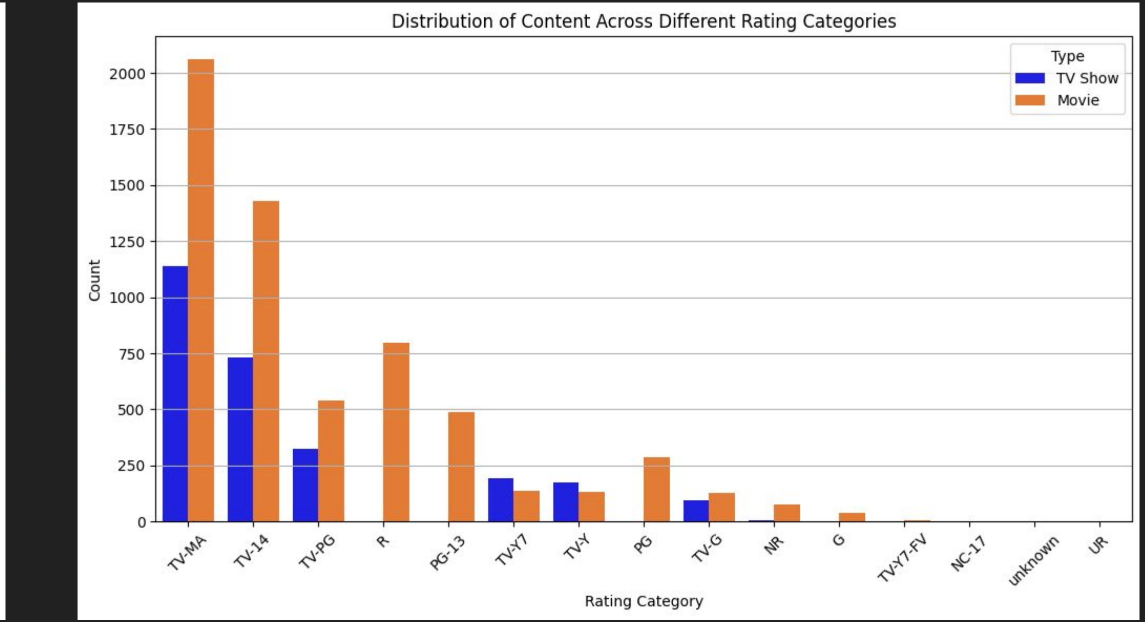

Top 10 Directors - Genre Associations
Director
Anurag Kashyap
Cathy Garcia-Molina
David Dhawan
Hakan Algül
Martin Scorsese
Rajiv Chilaka
Steven Spielberg
Suhas Kadav
Youssef Chahine
Yılmaz Erdoğan
Action & Adventure
Children & Family Movies
Classic Movies
Comedies
Crime TV Shows
Documentaries
Dramas
Horror Movies
Independent Movies
International Movies
International TV Shows
Music & Musicals
Romantic Movies
Sci-Fi & Fantasy
Sports Movies
TV Dramas
Thrillers
Genre
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5

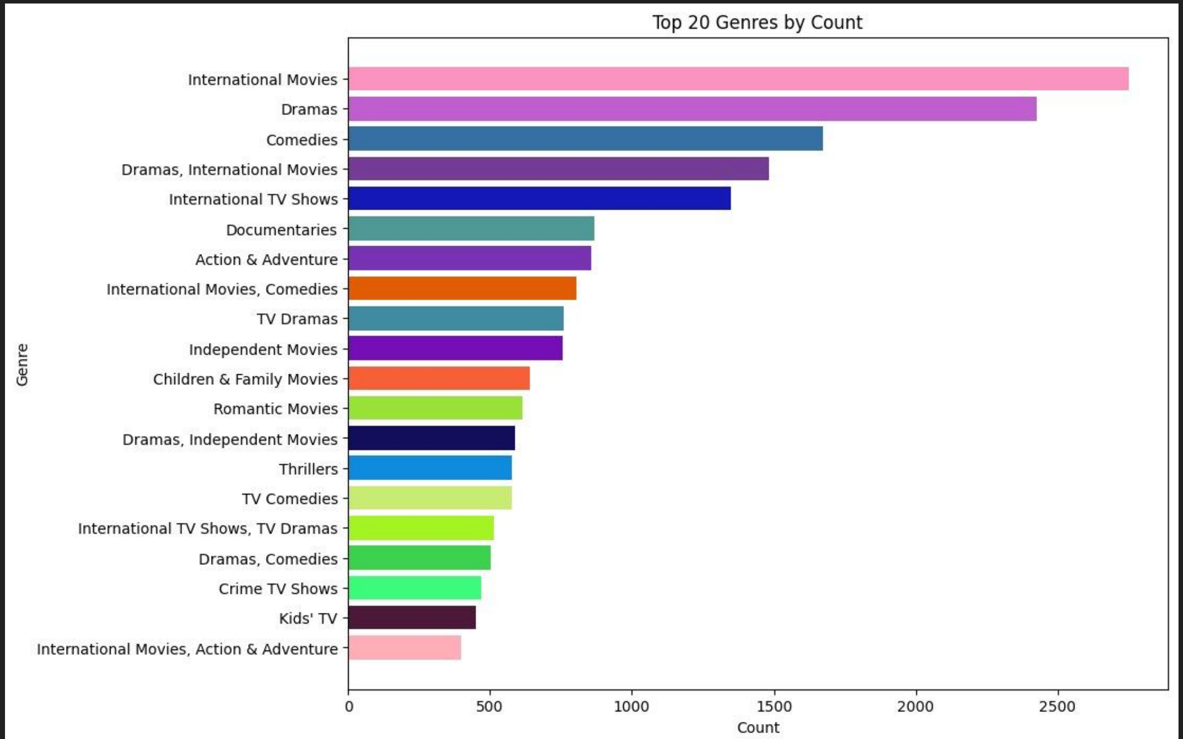
▶ Main Slide

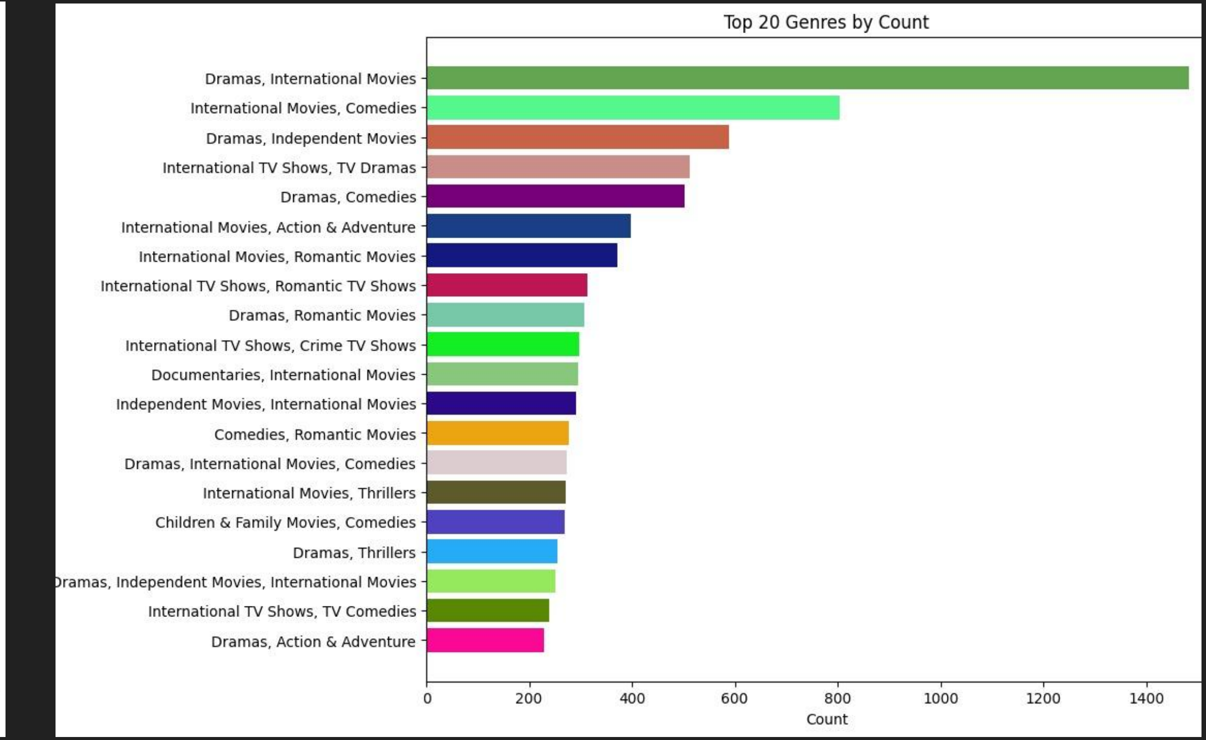

Comparison of Movies and TV Shows Count Each Year
Movies
TV Shows
Number of Content Added
0
100
200
300
400
500
600
700
800
1980
1990
2000
2010
2020
Year

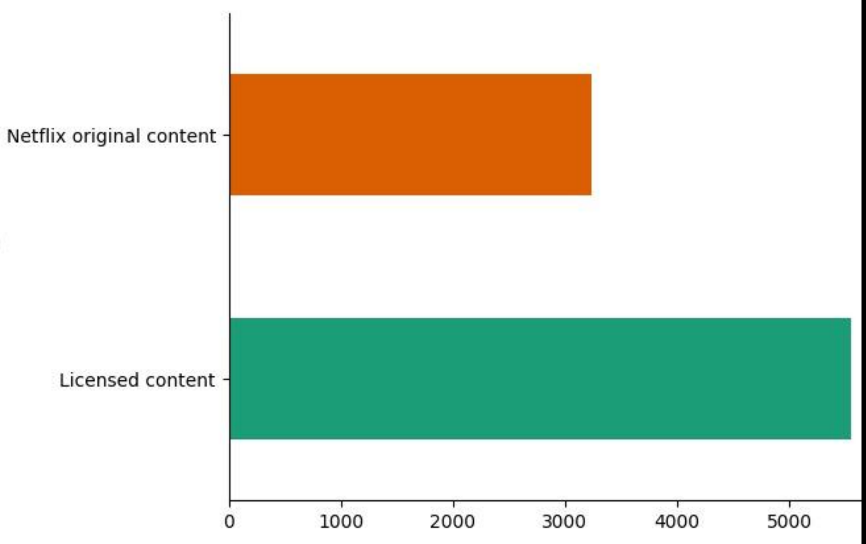

Genre and Country Diversity Over Time
Genre Diversity
Rating Diversity
Country Diversity
Number of Unique Items
0
10
20
30
40
50
60
70
80
1980
1985
1990
1995
2000
2005
2010
2015
2020
Release Year

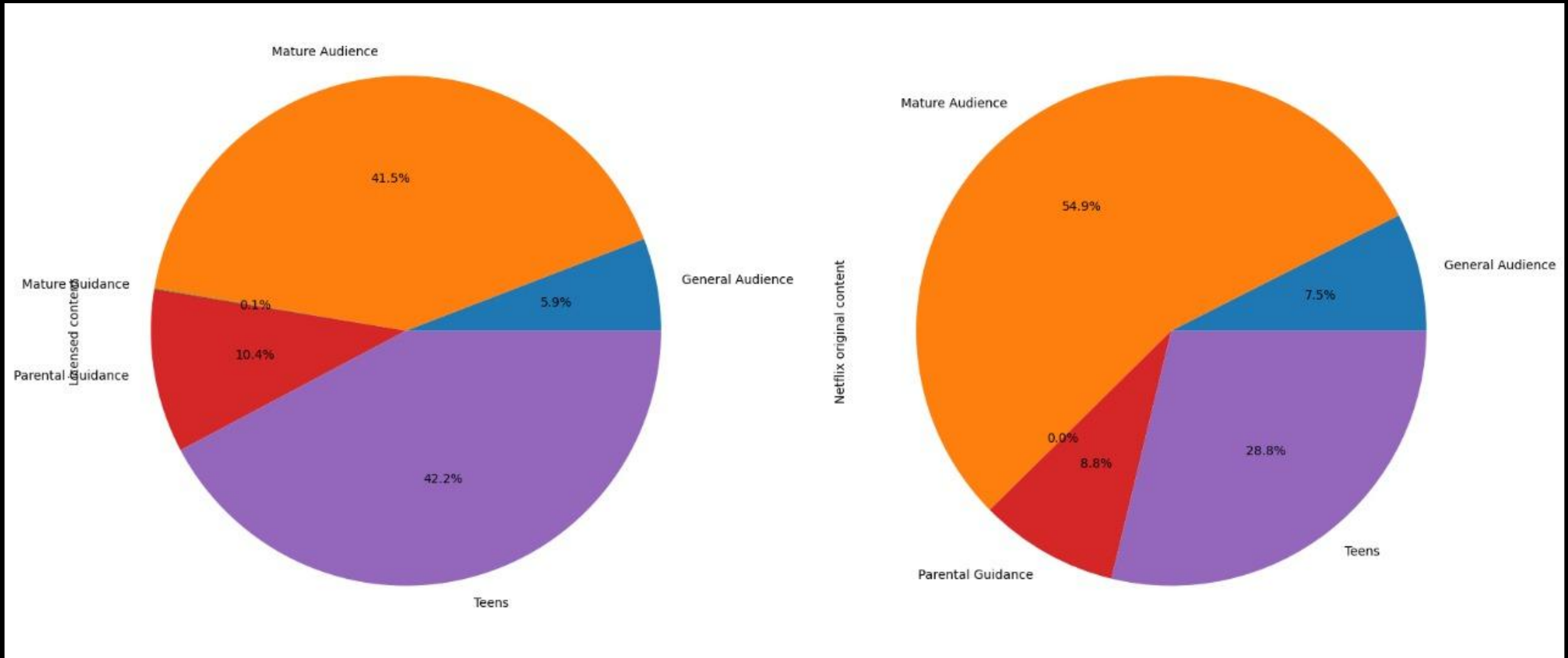

Distribution of Movie Durations by Release Year (After 1980)
Duration (minutes)
0
50
100
150
200
250
300
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Release Year

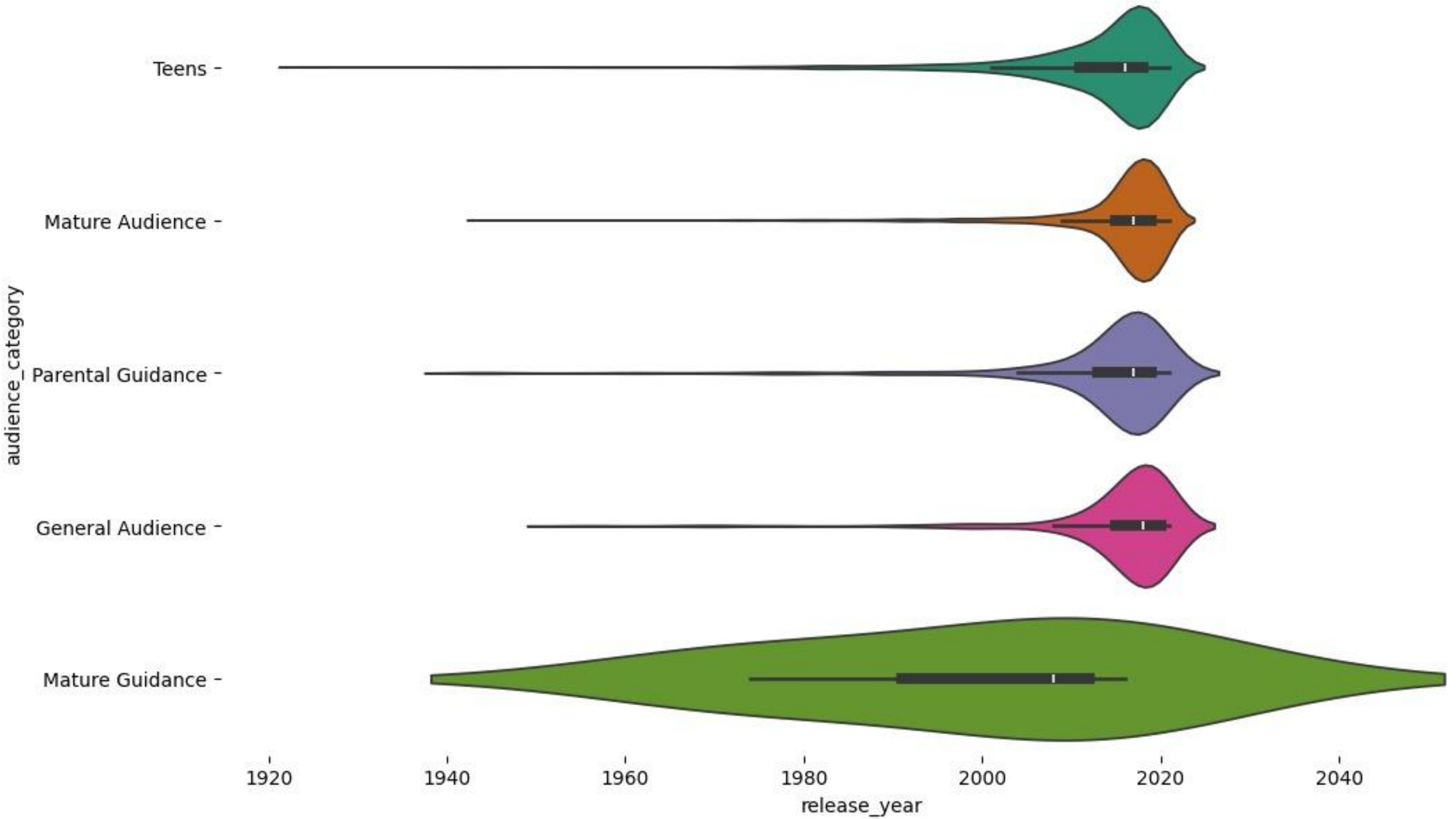

Distribution of Content Across Different Rating Categories
Type
TV Show
Movie
Count
0
250
500
750
1000
1250
1500
1750
2000
TV-MA
TV-14
TV-PG
R
PG-13
TV-Y7
TV-Y
PG
TV-G
NR
G
TV-Y7-FV
NC-17
unknown
UR
Rating Category


Top 20 Genres by Count
International Movies
Dramas
Comedies
Dramas, International Movies
International TV Shows
Documentaries
Action & Adventure
International Movies, Comedies
TV Dramas
Independent Movies
Children & Family Movies
Romantic Movies
Dramas, Independent Movies
Thrillers
TV Comedies
International TV Shows, TV Dramas
Dramas, Comedies
Crime TV Shows
Kids' TV
International Movies, Action & Adventure
Genre
0
500
1000
1500
2000
2500
Count


Top 20 Genres by Count
Dramas, International Movies
International Movies, Comedies
Dramas, Independent Movies
International TV Shows, TV Dramas
Dramas, Comedies
International Movies, Action & Adventure
International Movies, Romantic Movies
International TV Shows, Romantic TV Shows
Dramas, Romantic Movies
International TV Shows, Crime TV Shows
Documentaries, International Movies
Independent Movies, International Movies
Comedies, Romantic Movies
Dramas, International Movies, Comedies
International Movies, Thrillers
Children & Family Movies, Comedies
Dramas, Thrillers
Dramas, Independent Movies, International Movies
International TV Shows, TV Comedies
Dramas, Action & Adventure
0
200
400
600
800
1000
1200
1400
Count

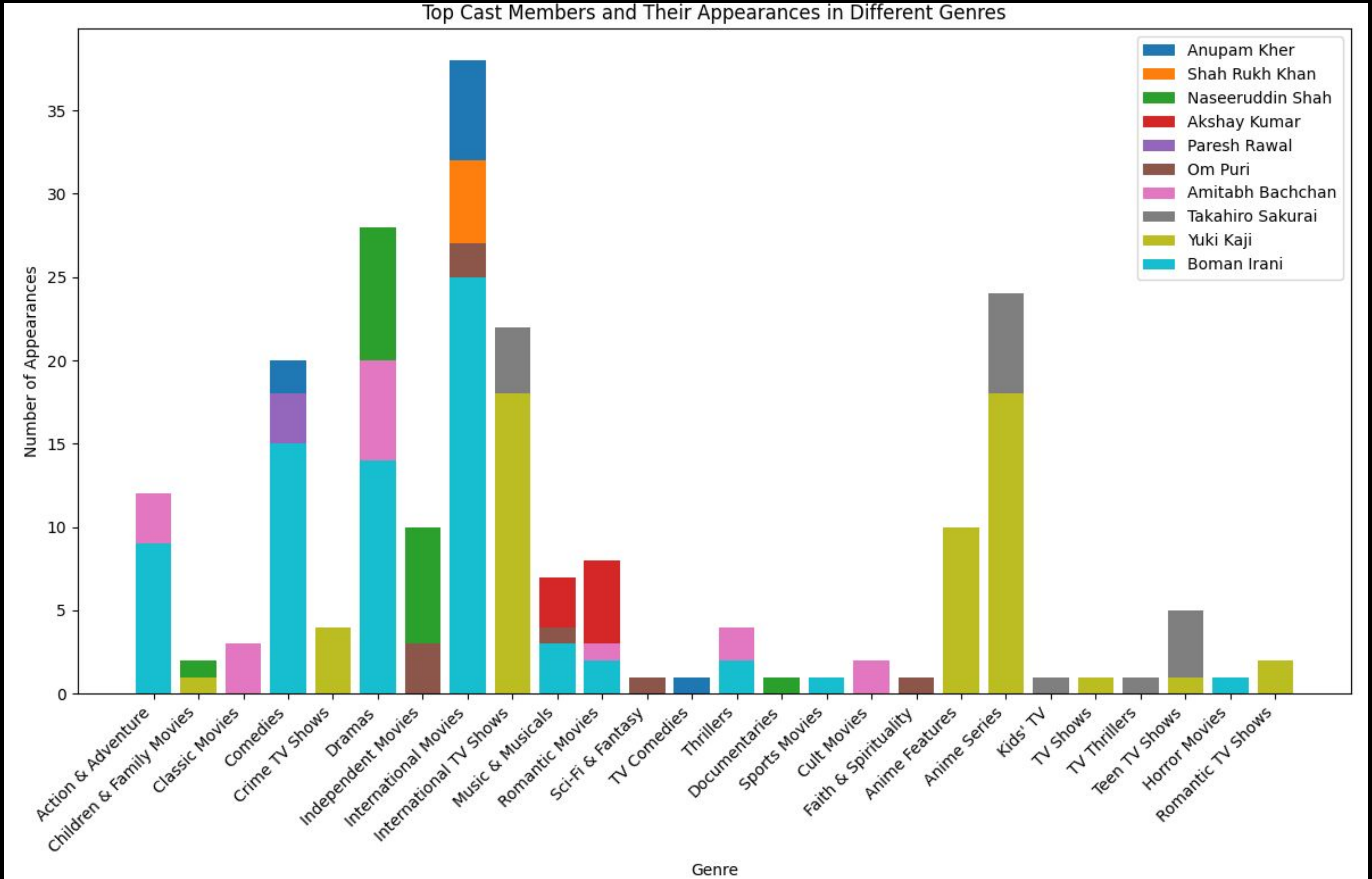

Netflix original content
Licensed content
0
1000
2000
3000
4000
5000


Mature Audience
41.5%
General Audience
5.9%
Mature Guidance
0.1%
Licensed content
Parental Guidance
10.4%
42.2%
Teens
Mature Audience
54.9%
General Audience
7.5%
Netflix original content
0.0%
8.8%
28.8%
Parental Guidance
Teens


Teens
Mature Audience
Parental Guidance
General Audience
Mature Guidance
audience_category
1920
1940
1960
1980
2000
2020
2040
release_year

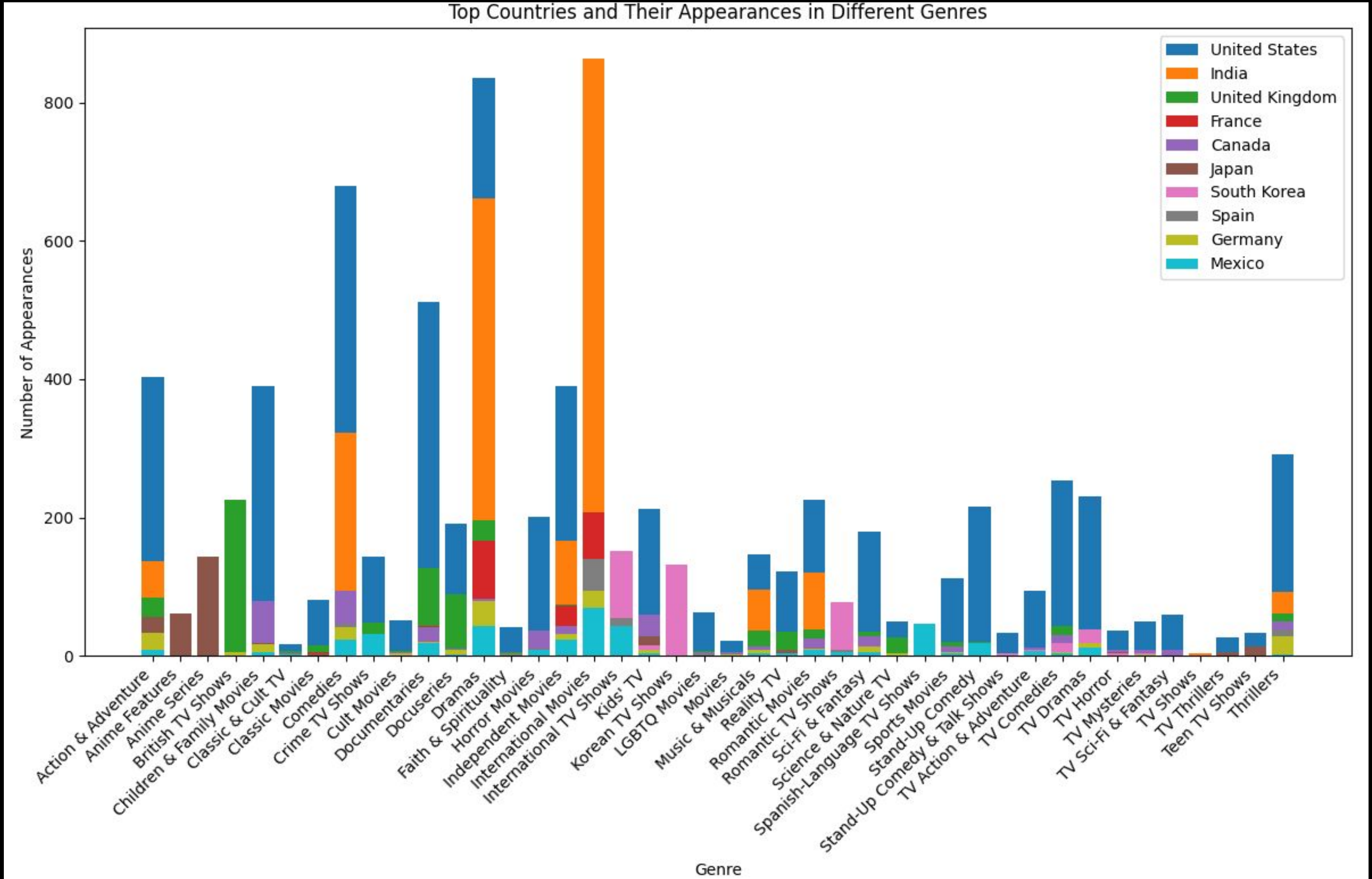
THANKS

GROUP 2



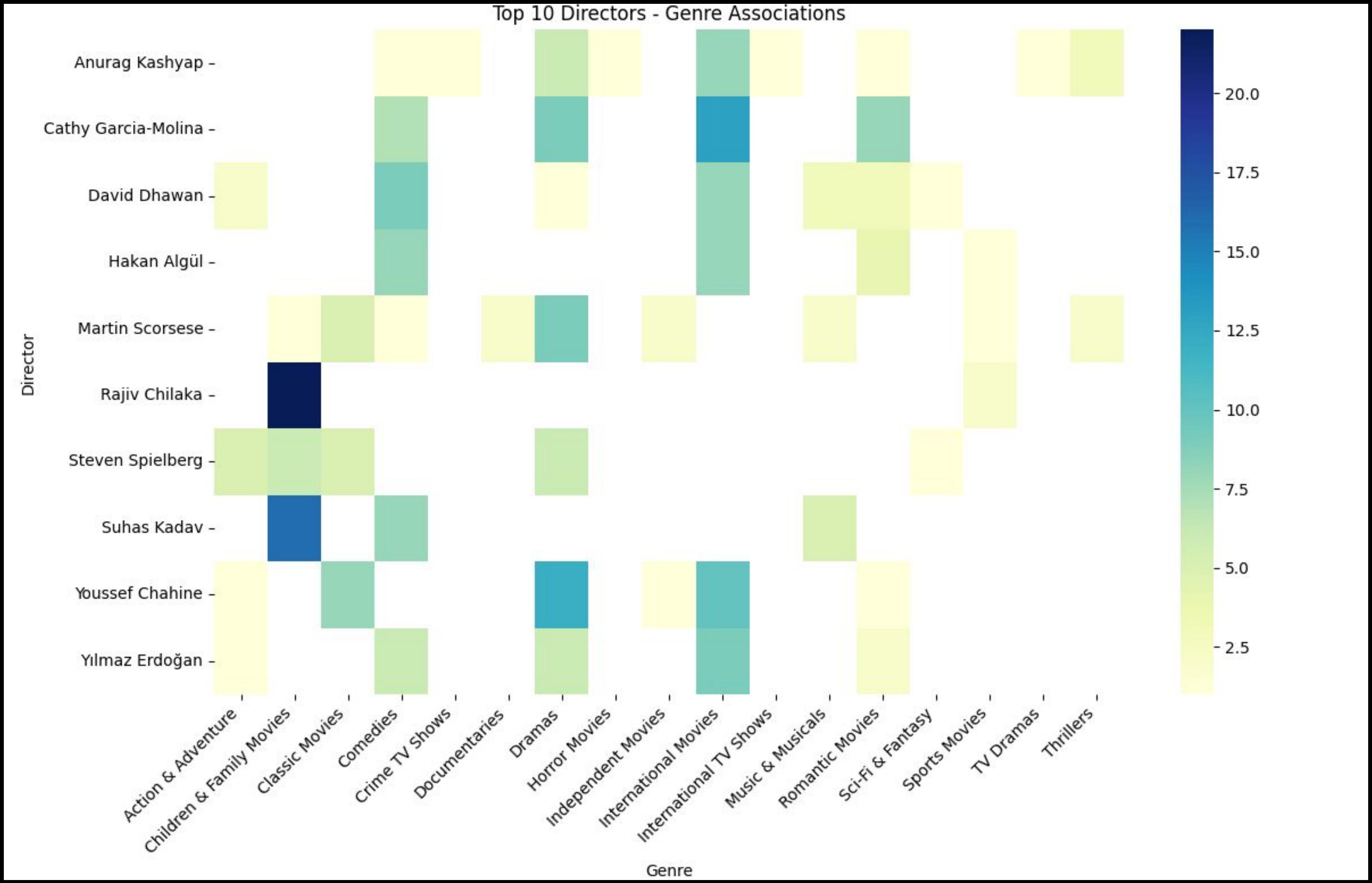
Anupam Kher	- 40
Shah Rukh Khan	- 34
Naseeruddin Shah	- 31
Akshay Kumar	- 29
Om Puri	- 29

▶ Main Slide



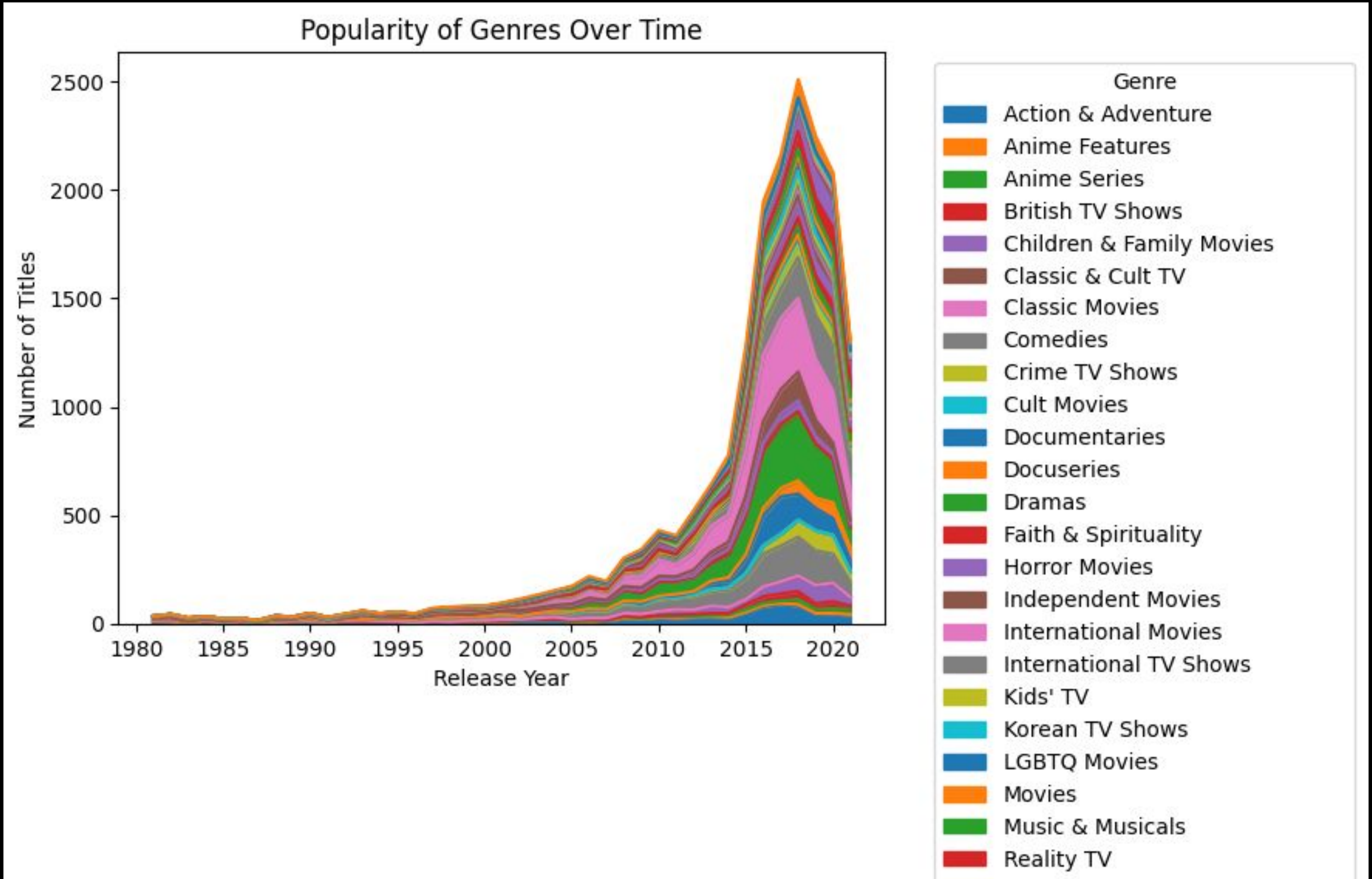
United States	6779
India	2804
United Kingdom	1779
France	916
Canada	877

▶ Main Slide

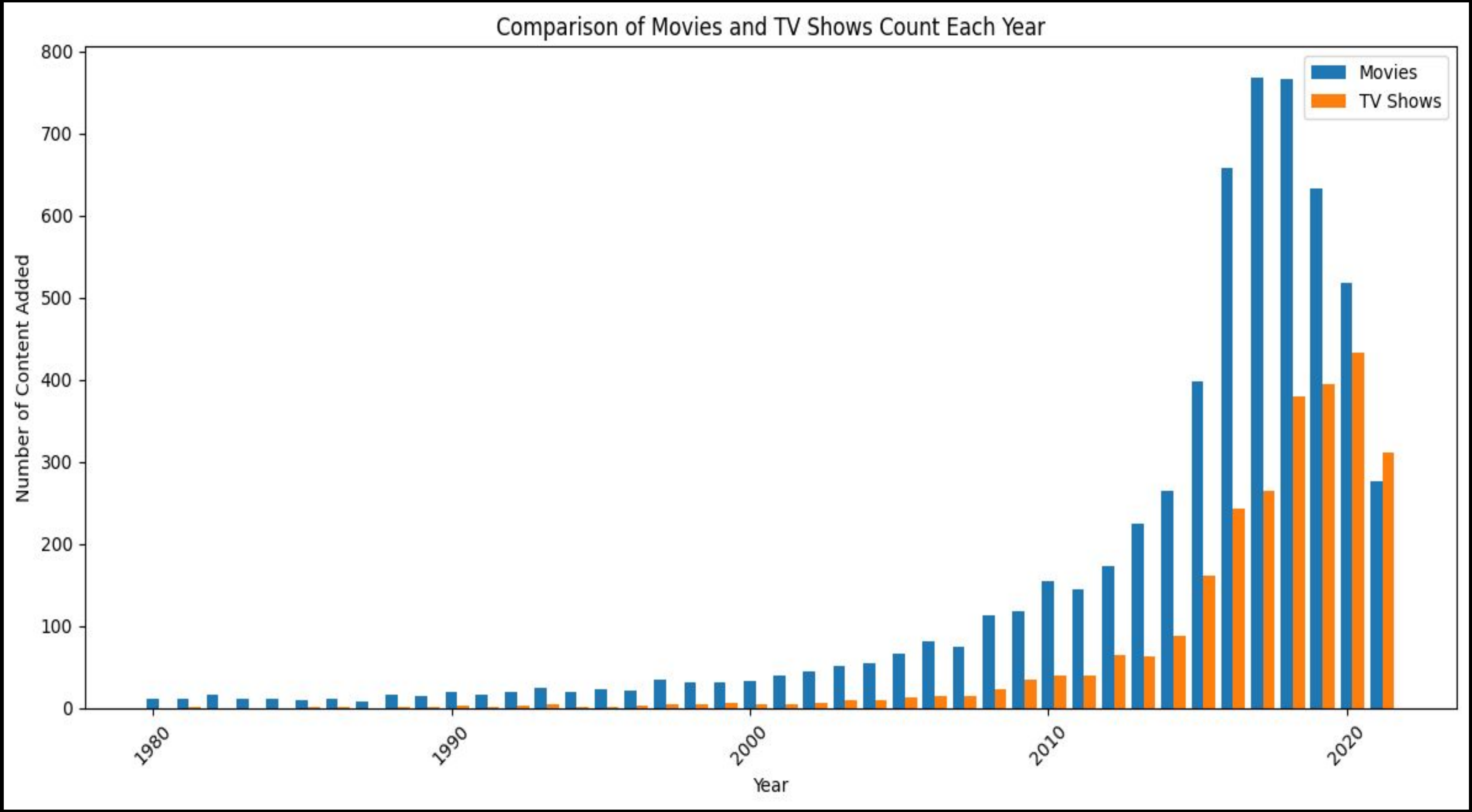


Rajiv Chilaka	19
Suhas Kadav	16

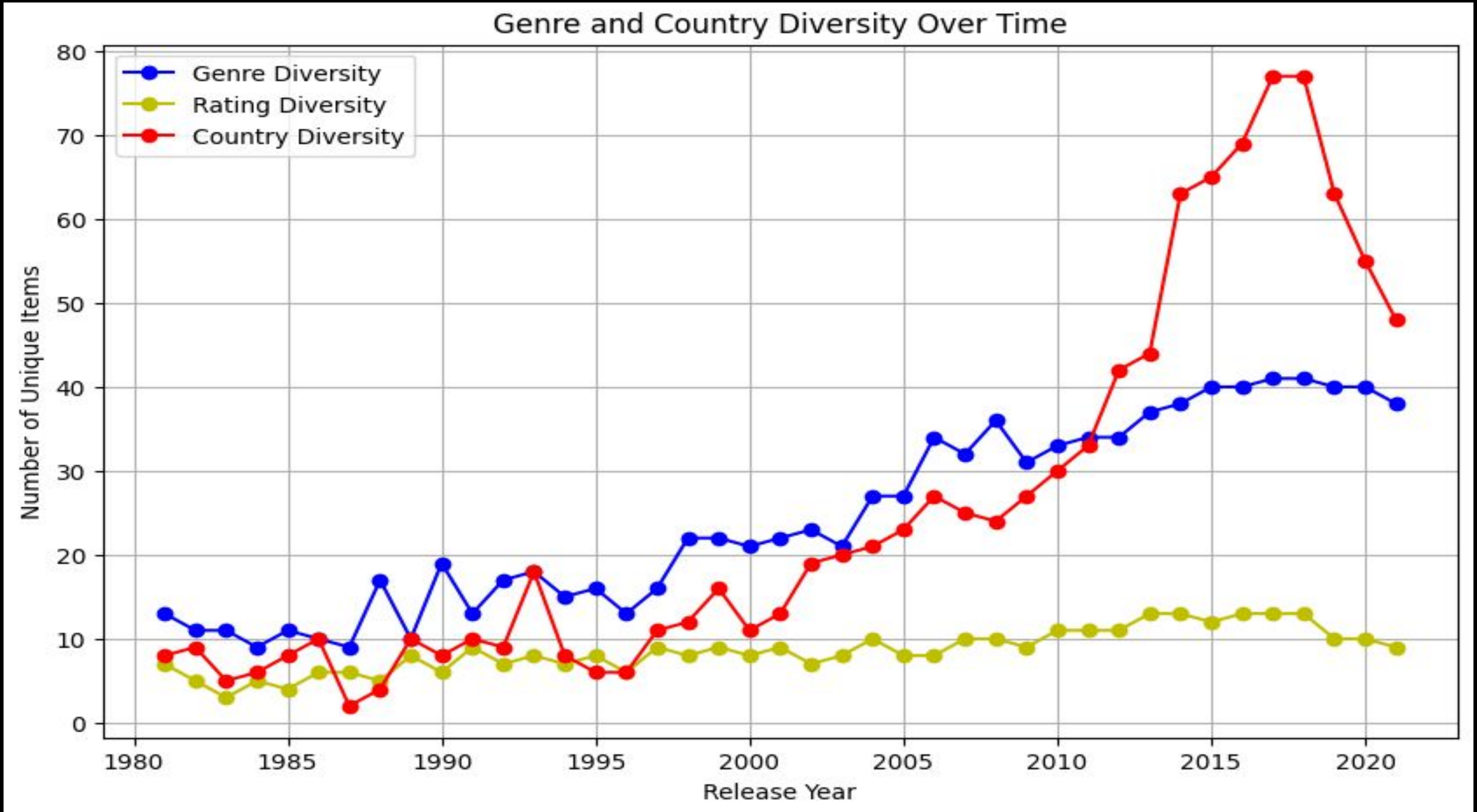
▶ Main Slide


Popularity of Genres Over Time
Number of Titles
2500
2000
1500
1000
500
0
1980
1985
1990
1995
2000
2005
2010
2015
2020
Release Year
Genre
Action & Adventure
Anime Features
Anime Series
British TV Shows
Children & Family Movies
Classic & Cult TV
Classic Movies
Comedies
Crime TV Shows
Cult Movies
Documentaries
Docuseries
Dramas
Faith & Spirituality
Horror Movies
Independent Movies
International Movies
International TV Shows
Kids' TV
Korean TV Shows
LGBTQ Movies
Movies
Music & Musicals
Reality TV

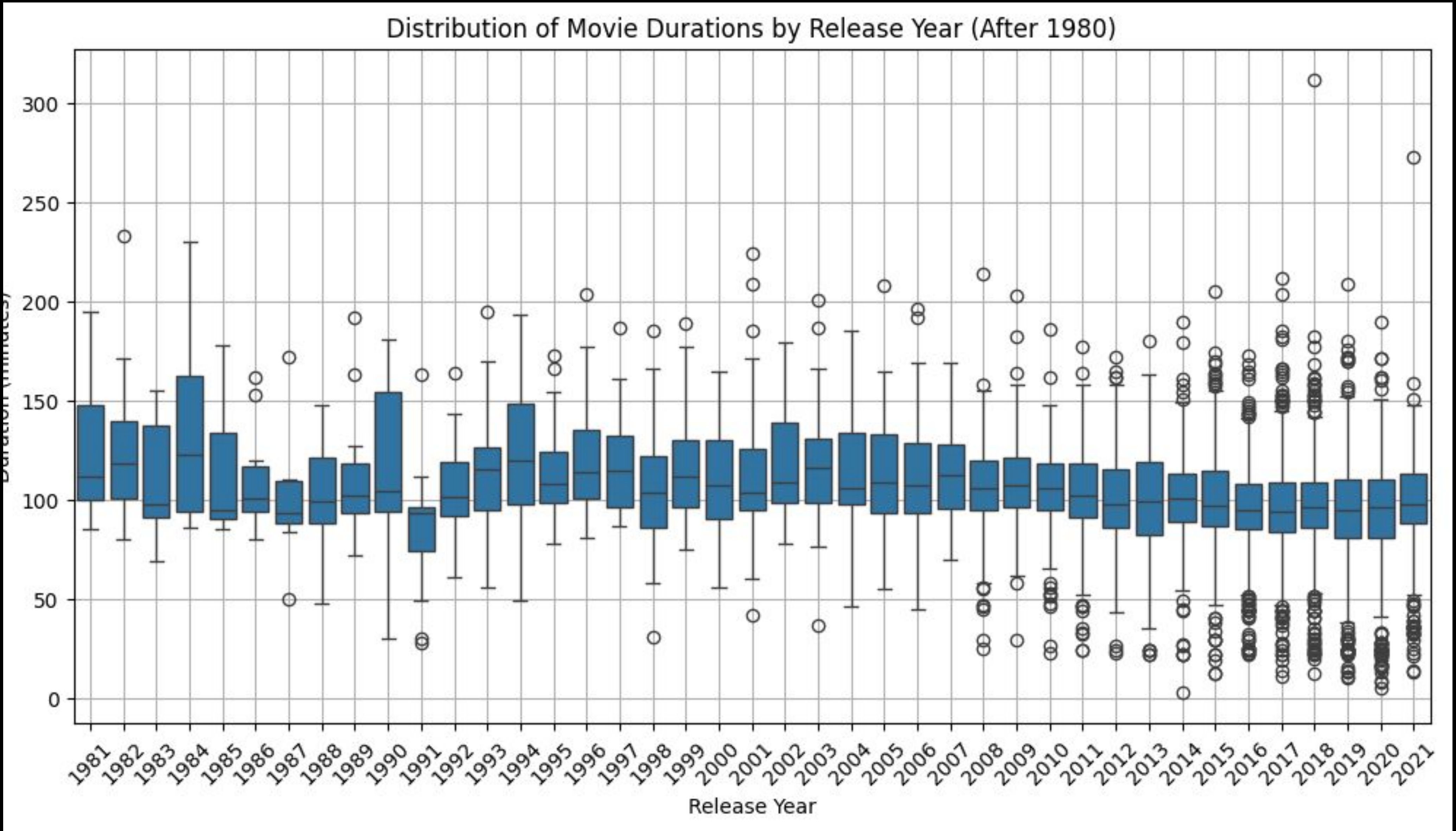
▶ Main Slide


Comparison of Movies and TV Shows Count Each Year
Movies
TV Shows
Number of Content Added
0
100
200
300
400
500
600
700
800
Year
1980
1990
2000
2010
2020

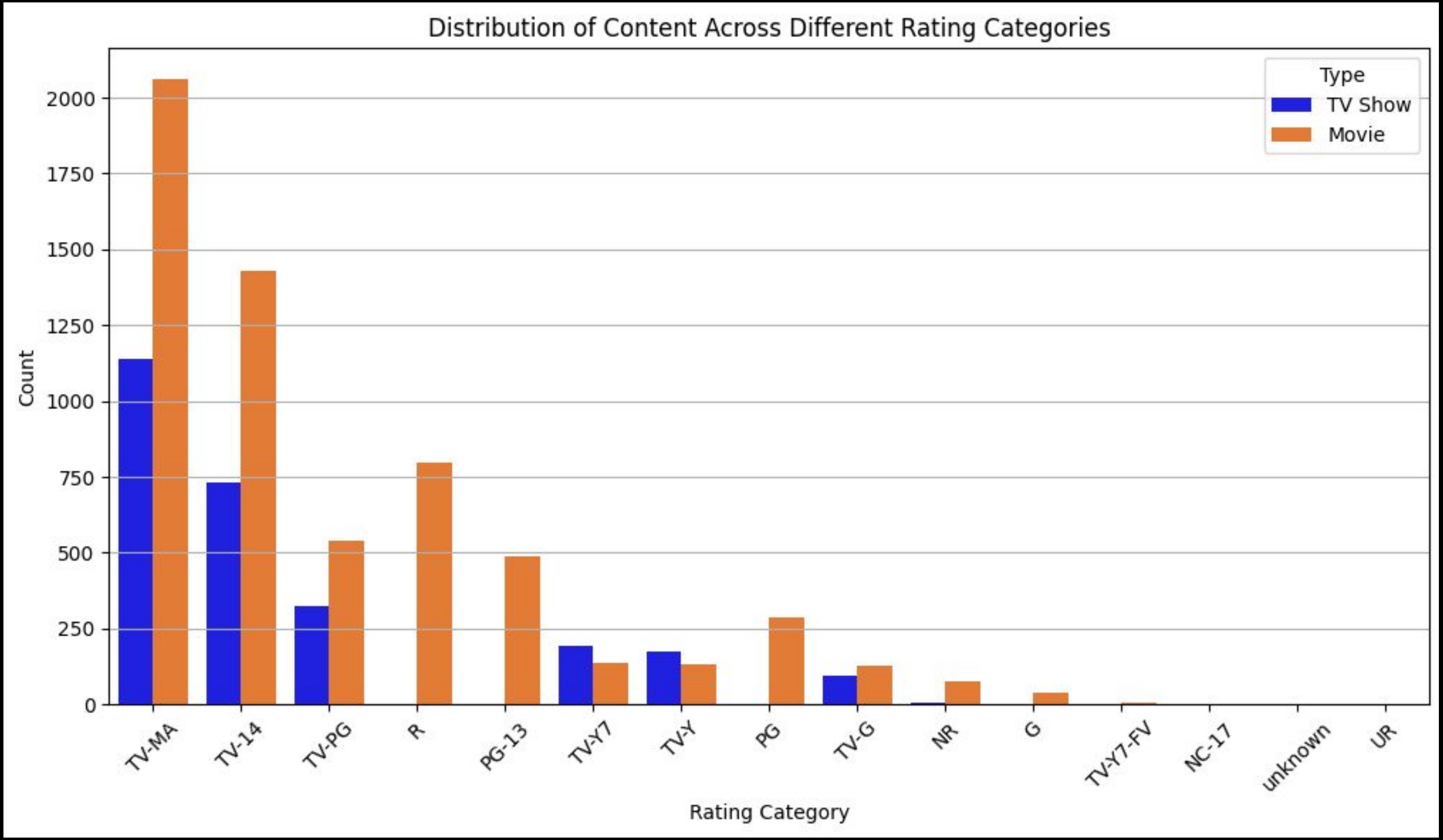
▶ Main Slide


Genre and Country Diversity Over Time
Genre Diversity
Rating Diversity
Country Diversity
Number of Unique Items
0
10
20
30
40
50
60
70
80
1980
1985
1990
1995
2000
2005
2010
2015
2020
Release Year

▶ Main Slide

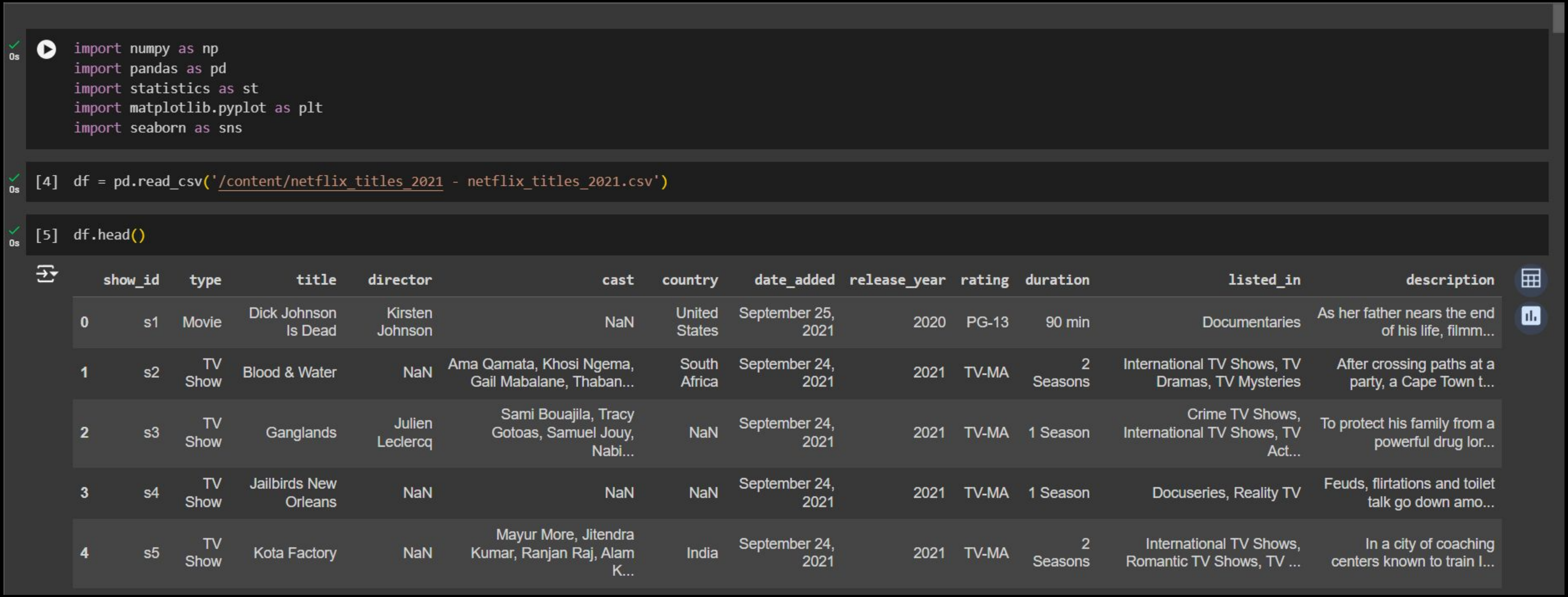

Distribution of Movie Durations by Release Year (After 1980)
Duration (minutes)
300
250
200
150
100
50
0
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Release Year

▶ Main Slide


Distribution of Content Across Different Rating Categories
Type
TV Show
Movie
Count
0
250
500
750
1000
1250
1500
1750
2000
TV-MA
TV-14
TV-PG
R
PG-13
TV-Y7
TV-Y
PG
TV-G
NR
G
TV-Y7-FV
NC-17
unknown
UR
Rating Category

▶ Main Slide

LOADING DATA



```
0s
import numpy as np
import pandas as pd
import statistics as st
import matplotlib.pyplot as plt
import seaborn as sns
```

```
0s
[4] df = pd.read_csv('/content/netflix_titles_2021 - netflix_titles_2021.csv')
```

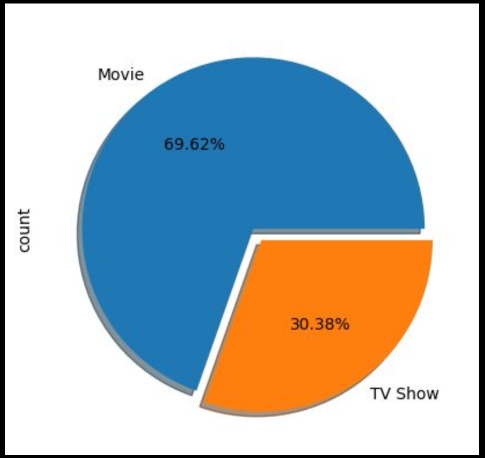
```
0s
[5] df.head()
```

	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...

Objective:

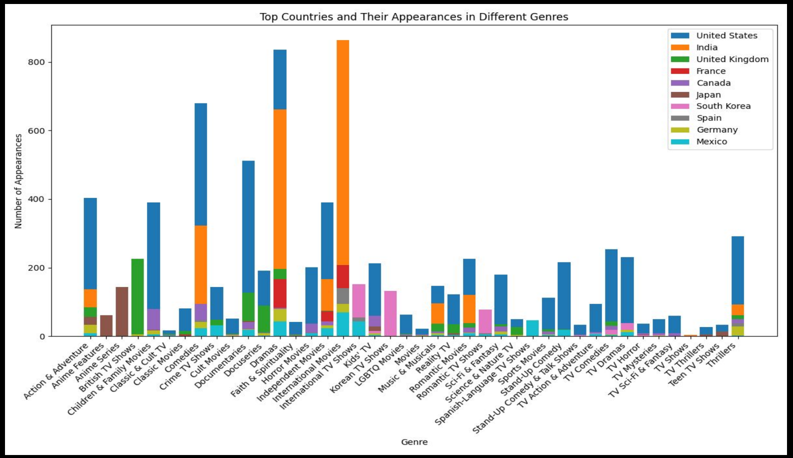
1. Content Analysis
2. Audience Engagement
3. Geographic Analysis

BASIC ANALYSIS



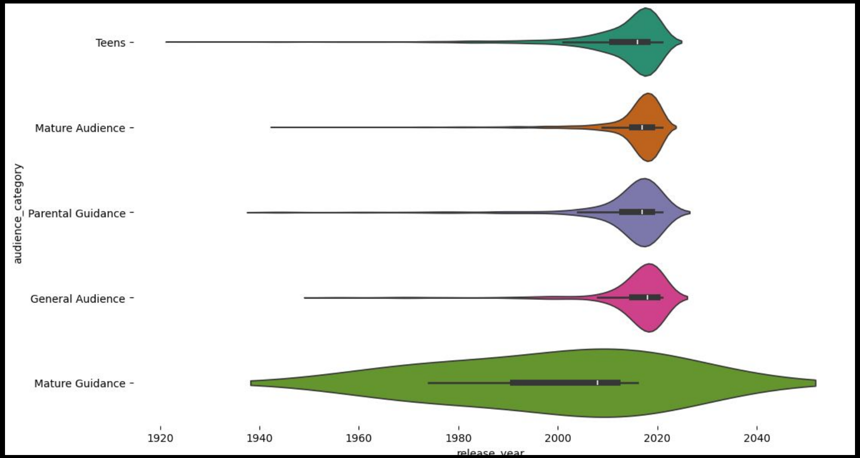
1

INSIGHTFUL ANALYSIS



2

IN DEPT ANALYSIS



3