```
# This Python 3 environment comes with many helpful analytics
libraries installed
# It is defined by the kaggle/python Docker image:
https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/"
directory
# For example, running this (by clicking run or pressing Shift+Enter)
will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/)
that gets preserved as output when you create a version using "Save &
Run All"
# You can also write temporary files to /kaggle/temp/, but they won't
be saved outside of the current session
```

```
/kaggle/input/netflix-shows/netflix_titles.csv
```

Data Cleaning & Preprocessing

- Removed or handled missing values (notably in director)
- Extracted year values from date columns
- Aggregated and grouped categorical variables
- Converted duration formats for analysis
- Removed inconsistencies for visualization clarity

Tools & Libraries Used

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- Plotly Express

```
df = pd.read_csv('/kaggle/input/netflix-shows/netflix_titles.csv')

# df overview
df.head()
```

```
  show_id     type                  title         director  \
0      s1    Movie   Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show          Blood & Water              NaN
```

```
2      s3  TV Show              Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans              NaN
4      s5  TV Show           Kota Factory              NaN

                                                   cast        country  \
0                                                   NaN  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...            NaN
3                                                   NaN            NaN
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...          India

           date_added  release_year rating   duration  \
0  September 25, 2021          2020  PG-13     90 min
1  September 24, 2021          2021  TV-MA  2 Seasons
2  September 24, 2021          2021  TV-MA   1 Season
3  September 24, 2021          2021  TV-MA   1 Season
4  September 24, 2021          2021  TV-MA  2 Seasons

                                        listed_in  \
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 ...

                                         description
0  As her father nears the end of his life, filmm...
1  After crossing paths at a party, a Cape Town t...
2  To protect his family from a powerful drug lor...
3  Feuds, flirtations and toilet talk go down amo...
4  In a city of coaching centers known to train I...
```

Data Assessing

- overview --> head(), sample(), data card, shape, validity issues, duplicacy check
- seeking information --> completeness
- seeking description --> accuracy issue, validity issue, completeness
- distribution --> numerical columns, categorical(imbalancing) --> feature enggg.., feature selection

Data Visualization

- import major libraries
- univariate analysis
- bivariate analysis
- multivariate analysis
- information
- insights

---> Report

---> Dashboarding

```
# Importing Major Libraries

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px

df.shape
```

```
(8807, 12)
```

```
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
```

```
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
```

```
df.isnull().mean()*100
```

```
show_id          0.000000
type             0.000000
title            0.000000
director        29.908028
cast             9.367549
country          9.435676
date_added       0.113546
release_year     0.000000
rating           0.045418
duration         0.034064
listed_in        0.000000
description      0.000000
dtype: float64
```

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

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6173 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       6173 non-null   object
 1   type          6173 non-null   object
 2   title         6173 non-null   object
 3   director      6173 non-null   object
 4   cast          5700 non-null   object
 5   country       5751 non-null   object
 6   date_added    6173 non-null   object
 7   release_year  6173 non-null   int64 
 8   rating        6172 non-null   object
 9   duration      6170 non-null   object
 10  listed_in     6173 non-null   object
 11  description   6173 non-null   object
dtypes: int64(1), object(11)
memory usage: 626.9+ KB
```

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

```
show_id          0
type             0
title            0
director         0
cast           473
country        422
```

```
date_added         0
release_year       0
rating             1
duration           3
listed_in          0
description        0
dtype: int64
```

```
df.dropna(subset=['country','rating','duration','cast'],inplace=True)
```

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

```
show_id          0
type             0
title            0
director         0
cast             0
country          0
date_added       0
release_year     0
rating           0
duration         0
listed_in        0
description      0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 5332 entries, 7 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       5332 non-null   object
 1   type          5332 non-null   object
 2   title         5332 non-null   object
 3   director      5332 non-null   object
 4   cast          5332 non-null   object
 5   country       5332 non-null   object
 6   date_added    5332 non-null   object
 7   release_year  5332 non-null   int64 
 8   rating        5332 non-null   object
 9   duration      5332 non-null   object
 10  listed_in     5332 non-null   object
 11  description   5332 non-null   object
dtypes: int64(1), object(11)
memory usage: 541.5+ KB
```

```
df.describe()
```

```
       release_year
count   5332.000000
mean    2012.742123
std        9.625831
min     1942.000000
25%     2011.000000
50%     2016.000000
75%     2018.000000
max     2021.000000
```

```
df
```

```
     show_id     type                            title
director  \
7         s8    Movie                          Sankofa          Haile
Gerima
8         s9  TV Show  The Great British Baking Show      Andy
Devonshire
9        s10    Movie                      The Starling       Theodore
Melfi
12       s13    Movie                      Je Suis Karl  Christian
Schwochow
24       s25    Movie                             Jeans           S.
Shankar
...      ...      ...                               ...
...
8801   s8802    Movie                          Zinzana      Majid Al
Ansari
8802   s8803    Movie                           Zodiac        David
Fincher
8804   s8805    Movie                       Zombieland      Ruben
Fleischer
8805   s8806    Movie                             Zoom          Peter
Hewitt
8806   s8807    Movie                           Zubaan           Mozez
Singh

                                                   cast  \
7     Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
8     Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...
9     Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
12    Luna Wedler, Jannis Niewöhner, Milan Peschel, ...
24    Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...
...                                                 ...
8801  Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...
8802  Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
8804  Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
8805  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
8806  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...
```

```
                                                  country
date_added  \
7     United States, Ghana, Burkina Faso, United Kin...  September 24,
2021
8                                           United Kingdom  September 24,
2021
9                                            United States  September 24,
2021
12                               Germany, Czech Republic  September 23,
2021
24                                                 India  September 21,
2021
...                                                  ...
...
8801                        United Arab Emirates, Jordan       March 9,
2016
8802                                         United States   November 20,
2019
8804                                         United States    November 1,
2019
8805                                         United States    January 11,
2020
8806                                               India       March 2,
2019

      release_year rating   duration  \
7             1993  TV-MA    125 min
8             2021  TV-14  9 Seasons
9             2021  PG-13    104 min
12            2021  TV-MA    127 min
24            1998  TV-14    166 min
...            ...    ...        ...
8801          2015  TV-MA     96 min
8802          2007      R    158 min
8804          2009      R     88 min
8805          2006     PG     88 min
8806          2015  TV-14    111 min

                                             listed_in  \
7     Dramas, Independent Movies, International Movies
8                         British TV Shows, Reality TV
9                                     Comedies, Dramas
12                        Dramas, International Movies
24     Comedies, International Movies, Romantic Movies
...                                                ...
8801          Dramas, International Movies, Thrillers
8802                   Cult Movies, Dramas, Thrillers
8804                         Comedies, Horror Movies
8805               Children & Family Movies, Comedies
```

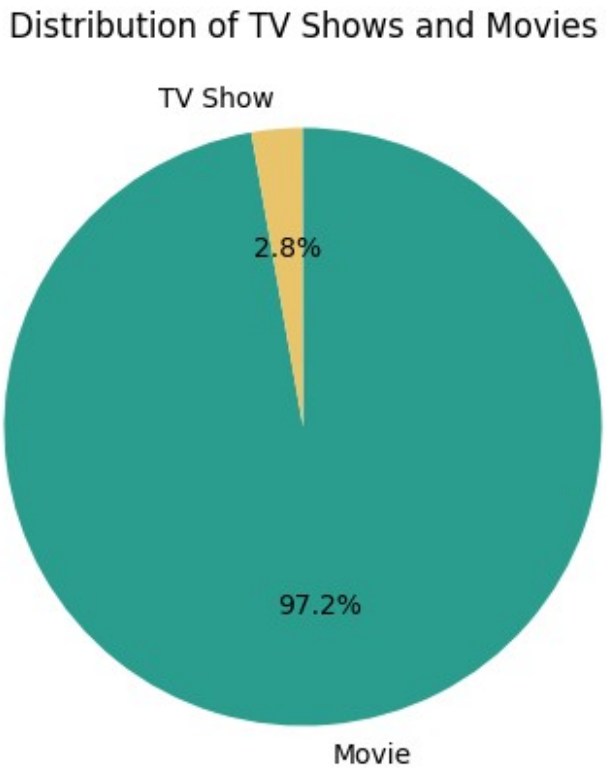
```
8806    Dramas, International Movies, Music & Musicals

                                                description
7     On a photo shoot in Ghana, an American model s...
8     A talented batch of amateur bakers face off in...
9     A woman adjusting to life after a loss contend...
12    After most of her family is murdered in a terr...
24    When the father of the man she loves insists t...
...                                                 ...
8801  Recovering alcoholic Talal wakes up inside a s...
8802  A political cartoonist, a crime reporter and a...
8804  Looking to survive in a world taken over by zo...
8805  Dragged from civilian life, a former superhero...
8806  A scrappy but poor boy worms his way into a ty...

[5332 rows x 12 columns]
```

Distribution of Type Content in Netflix

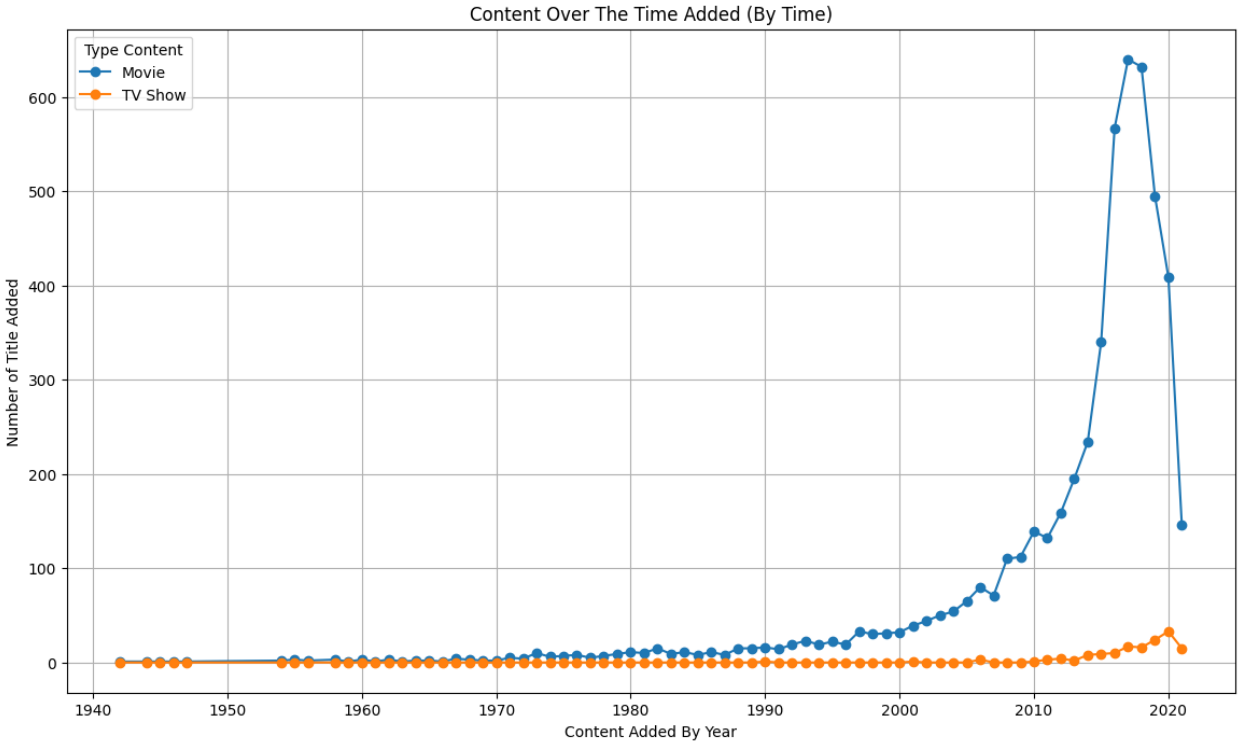
```
plt.figure(figsize=(12,5))
type_count = df['type'].value_counts()
plt.pie(type_count, labels=type_count.index, autopct='%1.1f%
%',startangle=100, colors=['#2A9D8F','#E9C46A'])
plt.title('Distribution of TV Shows and Movies')
plt.show()
```



Content To Be Added Over The Time

```
plt.figure(figsize=(14,8))
over_time =
df.groupby(['release_year','type']).size().unstack().fillna(0)
over_time.plot(kind='line', marker='o', figsize=(14, 8))
plt.title('Content Over The Time Added (By Time)')
plt.xlabel('Content Added By Year')
plt.ylabel('Number of Title Added')
plt.legend(title='Type Content')
plt.grid(True)
plt.show()
```

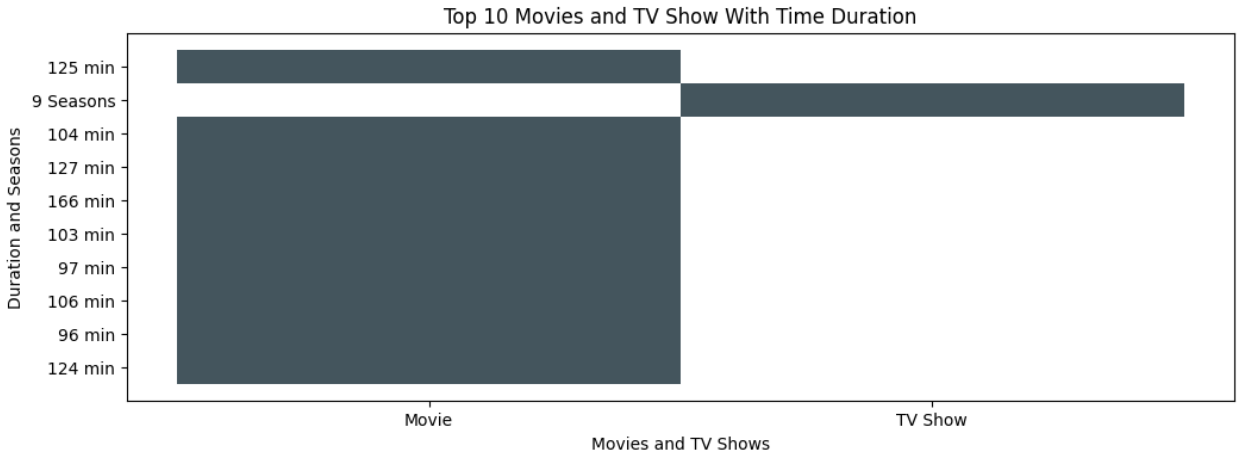
```
<Figure size 1400x800 with 0 Axes>
```




```
movie = df[df['type'] == 'Movie'].copy()
```

Top 10 Type Of Content With Duration and Seasons

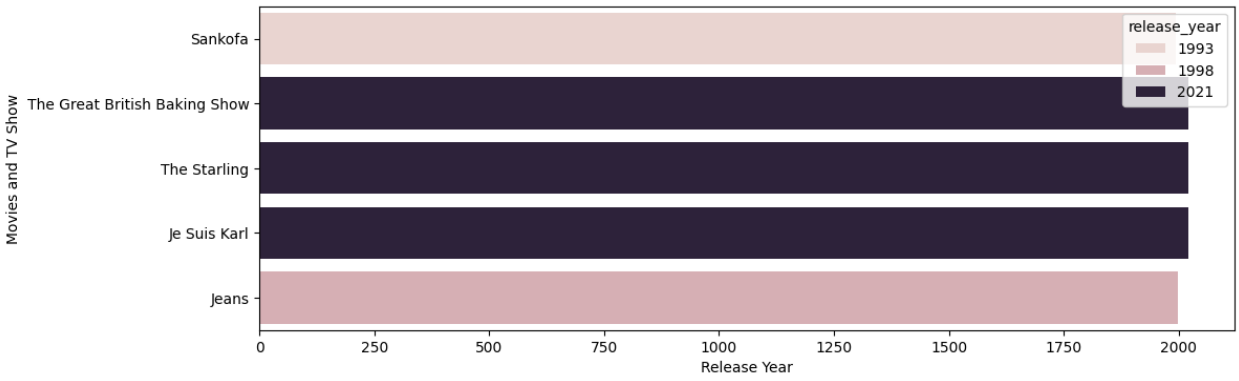
```
plt.figure(figsize=(12,4))
sns.histplot(data=top10,x='type',y='duration', color='skyblue')
plt.title('Top 10 Movies and TV Show With Time Duration')
plt.xlabel('Movies and TV Shows')
plt.ylabel('Duration and Seasons')
plt.show()
```



Top 5 Movies and TV Show With Release Year

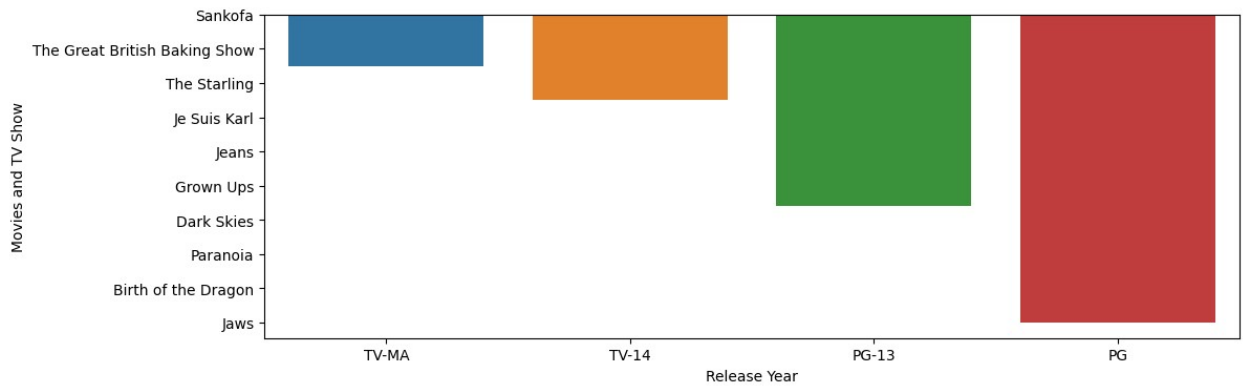
```
top5 = df.head(5)

plt.figure(figsize=(12,4))
sns.barplot(data=top5,y='title',x='release_year',hue='release_year')
plt.xlabel('Release Year')
plt.ylabel('Movies and TV Show')
plt.show()
```



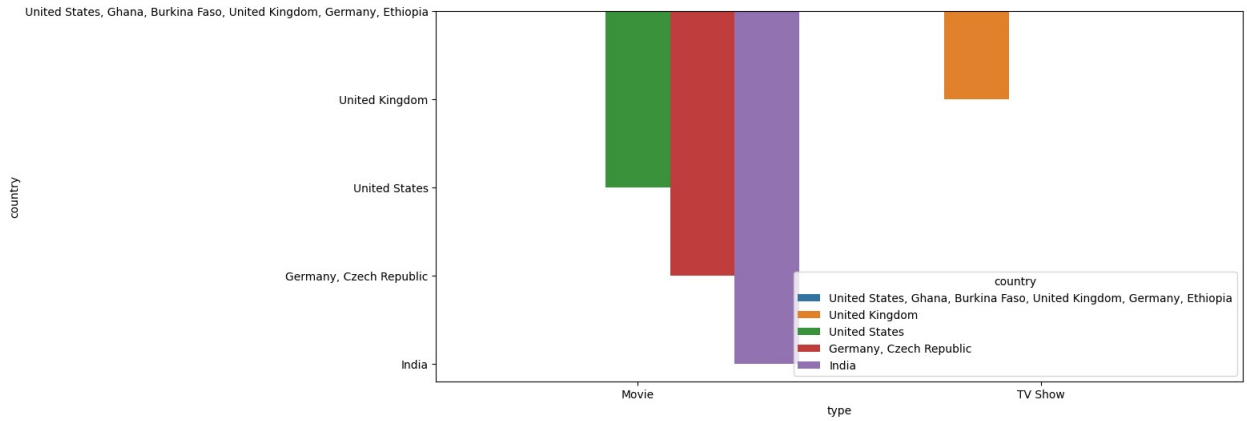
Top 10 Movies and TV Shows With Rating

```
plt.figure(figsize=(12,4))
sns.barplot(data=top10,y='title',x='rating',hue='rating',errorbar=None
)
plt.xlabel('Release Year')
plt.ylabel('Movies and TV Show')
plt.show()
```



Top 10 Country Producing Content on Netflix

```
plt.figure(figsize=(13,6))
sns.barplot(data=top5,x='type',y='country',hue='country')
plt.show()
```



Key Insights from the Analysis

1. Movies Dominate the Netflix Catalog

- Movies significantly outnumber TV Shows.
- Netflix appears to prioritize short-form content over long-running series.

1. Rapid Content Growth After 2015

- A sharp increase in content additions is observed post-2015.
- This aligns with Netflix's global expansion and aggressive original content strategy.

1. United States Leads Content Production

- The U.S. is the largest contributor, followed by India and the UK.
- Indicates Netflix's strong Western market focus, with growing international presence.

1. Mature Content Is Most Common

- Ratings such as TV-MA and TV-14 dominate.
- Netflix primarily targets a mature and young-adult audience.

1. Movie Durations Cluster Around 90–120 Minutes

- Most movies fall within standard feature-length runtimes.
- TV shows commonly have 1–3 seasons, indicating limited-series popularity.

Conclusion

This Netflix EDA project provides valuable insights into the platform's content strategy and catalog composition. The analysis reveals that Netflix heavily emphasizes movies, mature-rated content, and U.S.-produced titles, while simultaneously expanding its international footprint.

The surge in releases after 2015 highlights Netflix's transformation from a content distributor to a global content creator. Despite some metadata limitations, the dataset effectively supports trend analysis and strategic insights.

Overall, this project demonstrates how exploratory data analysis can uncover meaningful patterns in streaming platforms and can serve as a strong foundation for: