

Netflix Case Study

```
In [1]: """
The dataset provided to you consists of a list of all the TV shows/movies available on Netflix:

Show_id: Unique ID for every Movie / TV Show
Type: Identifier - A Movie or TV Show
Title: Title of the Movie / TV Show
Director: Director of the Movie
Cast: Actors involved in the movie/show
Country: Country where the movie/show was produced
Date_added: Date it was added on Netflix
Release_year: Actual Release year of the movie/show
Rating: TV Rating of the movie/show
Duration: Total Duration - in minutes or number of seasons
Listed_in: Genre
Description: The summary description
exploring a few questions:
What type of content is available in different countries?
How has the number of movies released per year changed over the last 20-30 years?
Comparison of tv shows vs. movies.
What is the best time to launch a TV show?
Analysis of actors/directors of different types of shows/movies.
Does Netflix has more focus on TV Shows than movies in recent years
Understanding what content is available in different countries

"""

Out[1]: """The dataset provided to you consists of a list of all the TV shows/movies available on Netflix.\n\nShow_id: Unique ID for every Movie / TV Show\nType: Identifier - A Movie or TV Show\nTitle: Title of the Movie / TV Show\nDirector: Director of the Movie\nCast: Actors involved in the movie/show\nCountry: Country where the movie/show was produced\nDate_added: Date it was added on Netflix\nRelease_year: Actual Release year of the movie/show\nRating: TV Rating of the movie/show\nDuration: Total Duration - in minutes or number of seasons\nListed_in: Genre\nDescription: The summary description\nexploring a few questions: \nWhat type of content is available in different countries?\nHow has the number of movies released per year changed over the last 20-30 years?\nComparison of tv show s vs. movies.\nWhat is the best time to launch a TV show?\n\nAnalysis of actors/directors of different types of shows/movies.\nDoes Netflix has more focus on TV Shows than movies in recent years\nComparison of tv show s vs. movies.\nWhat is the best time to launch a TV show?\n\nUnderstanding what content is available in different countries\n\n"""

In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

1. Defining Problem Statement and Analysing basic metrics

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

In [4]: df

Out[4]:
```

	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 Mabaleane, Thabani...	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, Nabil...	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	Jalbirds 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, Alan 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...
...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2021	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chahan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows x 12 columns

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

```
Out[5]:
```

	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 Mabaleane, 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, Nabil...	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	Jalbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down ento...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alan 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...

```
In [6]: df.tail()
```

```
Out[6]:
```

8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chahan...	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...

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

```
In [7]: df.shape
(8807, 12)

In [8]: df.columns
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in', 'description'], dtype='object')

In [9]: df.keys()
Out[9]:
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in', 'description'], dtype='object')

In [10]: categorical_columns = ['type', 'rating', 'listed_in']

In [11]: df[categorical_columns] = df[categorical_columns].astype('category')

In [12]: 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 Mabaleane, 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, Nabil...	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	Jalbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down ento...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alan 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...

```
In [13]: missing_values_count = df.isnull().sum()

In [14]: missing_values_count

Out[14]:
show_id      0
type         0
title        0
director    2634
cast        825
country     821
date_added   10
release_year  0
rating       4
duration     3
listed_in    0
description  0
dtype: int64
```

3. Non-Graphical Analysis: Value counts and unique attributes

```
In [15]: df["director"].unique()

Out[15]:
array(['Kirsten Johnson', nan, 'Julien Leclercq', ..., 'Majid Al Ansari', 'Peter Hewitt', 'Mozez Singh'], dtype=object)

In [16]: df["director"].value_counts()

Out[16]:
Rajiv Chilaka      19
Raul Campos, Jan Suter      18
Marcus Raboy      16
Suhass Kadav      16
Jay Karas      14
...
Raymie Muzquiz, Stu Livingston      1
Joe Menendez      1
Eric Bross      1
Will Eisenberg      1
Mozez Singh      1
Name: director, Length: 4528, dtype: int64

In [ ]:

What type of content is available in different countries?
```



```
In [40]: df['listed_in'] = df['listed_in'].str.split(',')
df = df.explode('listed_in')

In [42]: content_counts = df['listed_in'].value_counts()

In [43]: content_counts

Out[43]:
International Movies      3509
Dramas                   3199
Comedies                 1981
International TV Shows   1465
Action & Adventure       1182
Documentaries            1116
Independent Movies       1040
TV Dramas                852
Children & Family Movies  845
Thrillers                806
Romantic Movies         722
TV Comedies             639
Kids' TV                596
Crime TV Shows          541
Horror Movies           458
Docuseries              426
Music & Musicals         429
Romantic TV Shows       384
Sci-Fi & Fantasy         380
Stand-Up Comedy         344
British TV Shows        306
Reality TV Shows        279
Sports Movies           265
TV Action & Adventure    284
Spanish-Language TV Shows 262
Anime Series            192
Korean TV Shows         156
Classic Movies          151
LGBTQ Movies            116
TV Mysteries            113
Science & Nature TV     108
TV Sci-Fi & Fantasy      101
Cult Movies              93
Faith & Spirituality     89
TV Horror                86
Anime Features           78
Movies                   70
Teen TV Show             69
TV Thrillers             66
Stand-Up Comedy & Talk Shows 56
Classic & Cult TV        32
TV Shows                 16
Name: listed_in, dtype: int64

In [44]: df['listed_in'] = df['listed_in'].str.split(',')
df['country'] = df['country'].str.split(',')
df = df.explode('listed_in')
df = df.explode('country')
```

```
In [46]: content_by_country = df.groupby(['country', 'listed_in']).size().unstack(fill_value=0)

In [47]: content_by_country

Out[47]:
```

	listed_in	Action & Adventure	Anime Features	Anime Series	British TV Shows	Children & Family Movies	Classic & Cult TV	Classic Movies	Comedies	Crime TV Shows	Cult Movies	TV Action & Adventure	TV Action & Comedies	TV Dramas	TV Horror	TV Mysteries	TV Sci-Fi & Fantasy	TV Shows	TV Thrillers	Teen TV Shows	Thrill
Afghanistan		0	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	0	0
Albania		0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
Algeria		0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
Angola		1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
...	
Vatican City		0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
Venezuela		0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
Vietnam		2	0	0	0	0	0	0	0	2	0	...	0	0	0	0	0	0	0	0	0
West Germany		0	0	0	0	1	0	1	1	0	0	...	0	1	1	0	0	0	0	0	0
Zimbabwe		0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0

127 rows x 42 columns

How has the number of movies released over the last 20-30 years?

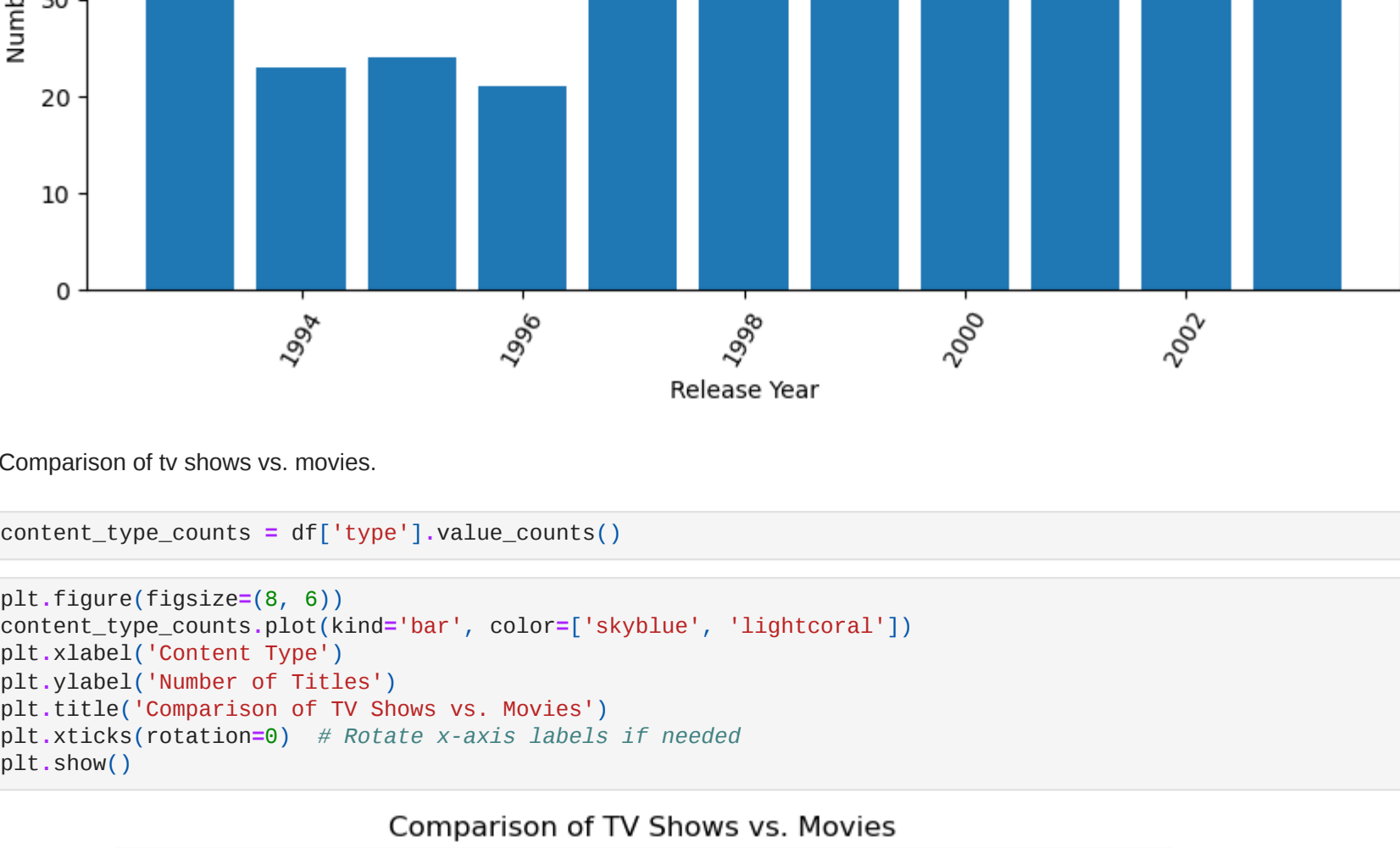
```
In [22]: movies_df = df[df['type'] == 'Movie']

In [23]: grouped = movies_df.groupby('release_year')

In [24]: movies_per_year = grouped.size().reset_index(name='Count')

In [25]: current_year = 2023
last_30_years = movies_per_year[
    (movies_per_year['release_year'] >= (current_year - 30)) &
    (movies_per_year['release_year'] <= (current_year - 20))
]

In [26]: plt.figure(figsize=(10, 6))
plt.bar(last_30_years['release_year'], last_30_years['Count'])
plt.xlabel('Release Year')
plt.ylabel('Number of Movies Released')
plt.title('Number of Movies Released per Year (Last 20-30 Years)')
plt.legend(['Best Time to Launch a TV Show'])
plt.xticks(rotation=60)
plt.show()
```



Comparison of tv shows vs. movies.

```
In [27]: content_type_counts = df['type'].value_counts()

In [28]: plt.figure(figsize=(8, 6))
content_type_counts.plot(kind='bar', color=['skyblue', 'lightcoral'])
plt.ylabel('Number of Titles')
plt.title('Comparison of TV Shows vs. Movies')
plt.xticks(rotation=0) # Rotate x-axis labels if needed
plt.show()
```

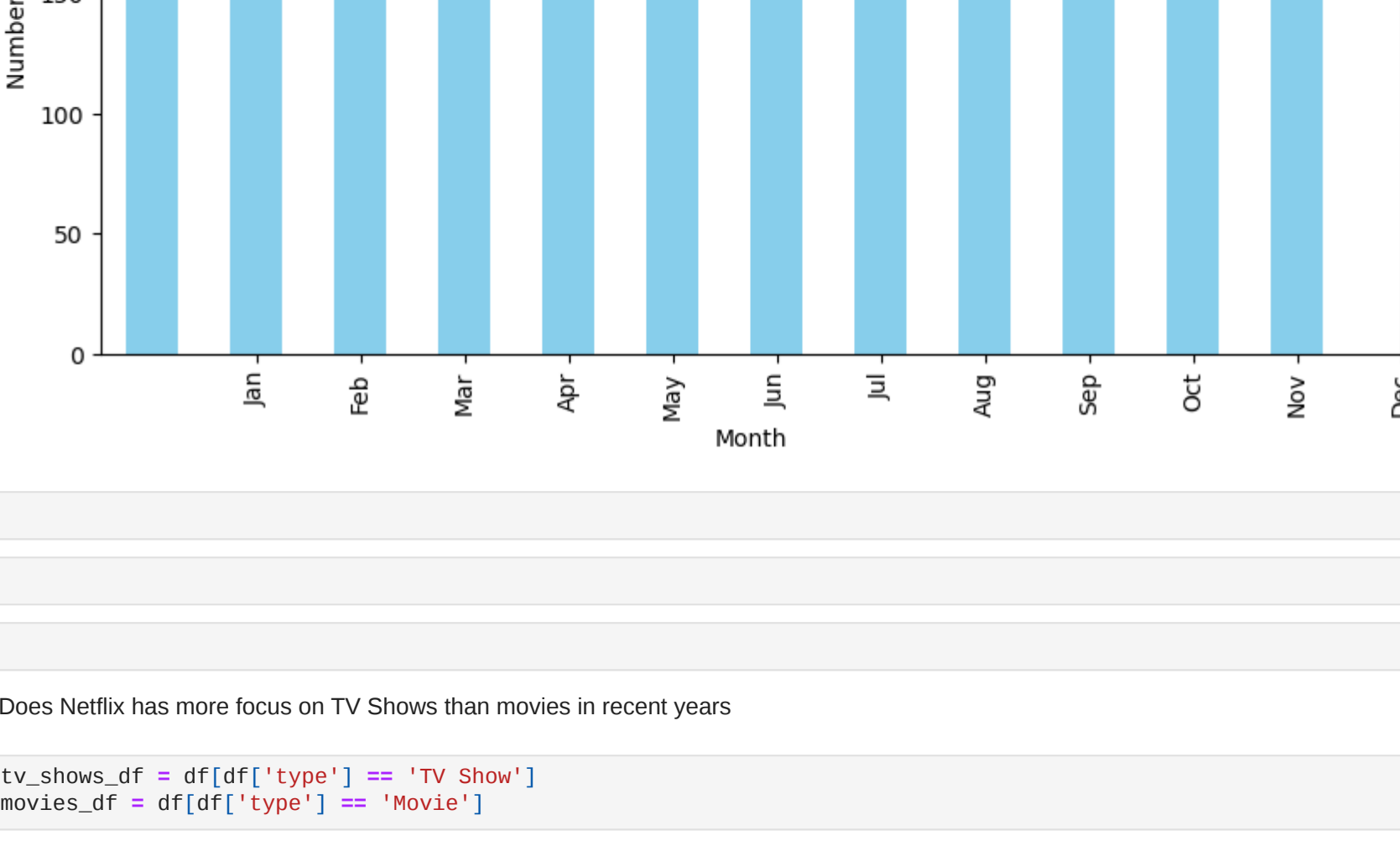
What is the best time to launch a TV show?

```
In [29]: tv_shows_df = df[df['type'] == 'TV Show']

In [ ]: tv_shows_df['Month_added'] = pd.to_datetime(tv_shows_df['date_added']).dt.month

In [31]: tv_show_counts_by_month = tv_shows_df['Month_added'].value_counts().sort_index()

In [32]: plt.figure(figsize=(10, 6))
tv_show_counts_by_month.plot(kind='bar', color='skyblue')
plt.ylabel('Month')
plt.xlabel('Number of TV Shows')
plt.title('Best Time to Launch a TV Show')
plt.xticks(range(1, 13), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])
plt.show()
```



```
In [ ]:

In [ ]:

In [ ]:
```

Does Netflix has more focus on TV Shows than movies in recent years

```
In [33]: tv_shows_df = df[df['type'] == 'TV Show']
movies_df = df[df['type'] == 'Movie']

In [34]: tv_shows_df['release_year'] = pd.to_datetime(tv_shows_df['date_added']).dt.year
movies_df['release_year'] = pd.to_datetime(movies_df['date_added']).dt.year

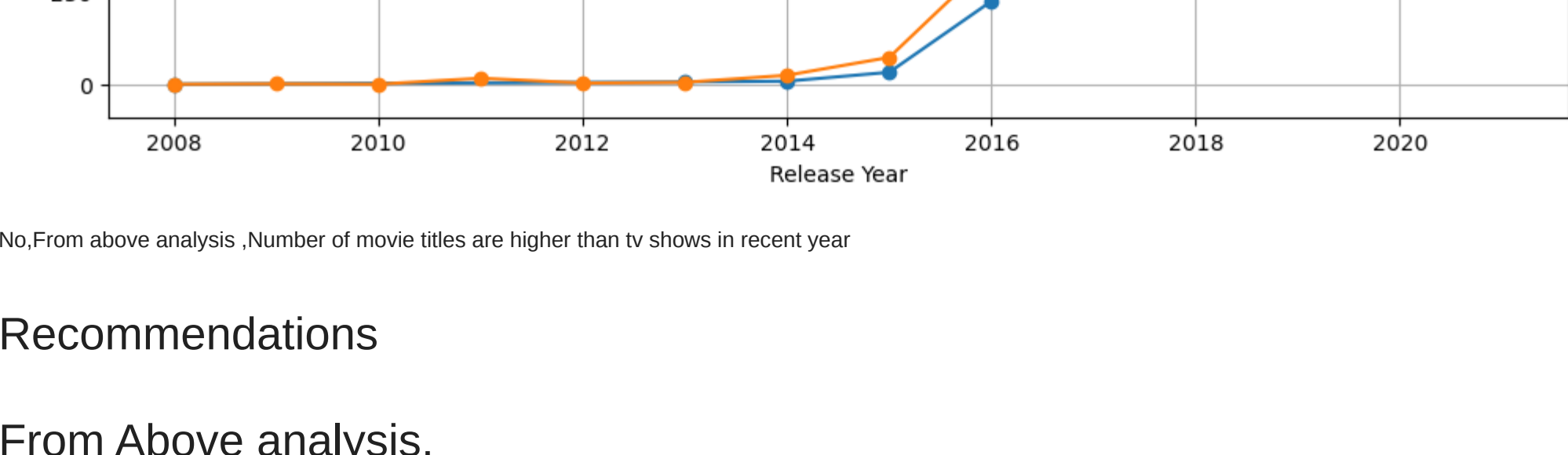
C:\Users\Akshay> satpute\AppData\LocalTemp\ipykernel_1612\322876281.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

C:\Users\Akshay> satpute\AppData\LocalTemp\ipykernel_1612\322876281.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
tv_shows_df['Release_year'] = pd.to_datetime(tv_shows_df['date_added']).dt.year
movies_df['Release_year'] = pd.to_datetime(movies_df['date_added']).dt.year

In [35]: movies_by_year = tv_shows_df.groupby('Release_year').size()
movies_by_year = movies_df.groupby('Release_year').size()

In [36]: plt.figure(figsize=(12, 6))
plt.plot(tv_shows_by_year.index, tv_shows_by_year, label='TV Shows', marker='o')
plt.plot(movies_by_year.index, movies_by_year, label='Movies', marker='o')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.title('Netflix Focus: TV Shows vs. Movies Over the Years')
plt.legend()
plt.grid(True)
plt.show()
```



No.From above analysis ,Number of movie titles are higher than tv shows in recent year

Recommendations

From Above analysis,

- 1.comparing to movies vs. Tv shows , number of movies are higher than number of tv shows
- 2.November is the best time to launch tvshow.
- 3.In last 20-30 years,Release of movies increased

```
In [ ]:
```