Brian_Reppeto_DSC550_Week_2

March 24, 2024

0.0.1 DSC 550 Week :

Activity 2.1

Author: Brian Reppeto 3/19/2024

```python
[101]: # import libraries

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

```python
[102]: # load the csv file

       net = pd.read_csv('netflix_titles.csv')
```

```python
[103]: # df shape

       net.shape
```

```
[103]: (8807, 12)
```

```python
[104]: # sort the df

       net.sort_values(by=['release_year'])
```

```
[104]:       show_id     type                                            title  \
       4250    s4251  TV Show                Pioneers: First Women Filmmakers*
       7790    s7791    Movie                                  Prelude to War
       8205    s8206    Movie                             The Battle of Midway
       8660    s8661    Movie  Undercover: How to Operate Behind Enemy Lines
       8739    s8740    Movie              Why We Fight: The Battle of Russia
       …         …        …                                                …
       862      s863  TV Show                                          Special
       861      s862    Movie          Little Singham Future mein Satakli
       857      s858  TV Show                                     The Neighbor
       876      s877  TV Show                                     Alma Matters
       965      s966  TV Show                                           Yasuke
```

```
                   director  \
4250                    NaN
7790            Frank Capra
8205             John Ford
8660             John Ford
8739  Frank Capra, Anatole Litvak
...                     ...
862                     NaN
861           Prakash Satam
857                     NaN
876                     NaN
965                     NaN


                                                   cast         country  \
4250                                                NaN             NaN
7790                                                NaN   United States
8205                      Henry Fonda, Jane Darwell   United States
8660                                                NaN   United States
8739                                                NaN   United States
...                                                 ...             ...
862   Ryan O'Connell, Jessica Hecht, Punam Patel, Ma…  United States
861   Anamaya Verma, Ganesh Divekar, Jigna Bharadhwa…             NaN
857   Quim Gutiérrez, Clara Lago, Adrián Pino, Catel…           Spain
876                                                NaN           India
965   LaKeith Stanfield, Takehiro Hira, Maya Tanida,…             NaN

             date_added  release_year rating   duration  \
4250  December 30, 2018          1925  TV-14   1 Season
7790      March 31, 2017         1942  TV-14     52 min
8205      March 31, 2017         1942  TV-14     18 min
8660      March 31, 2017         1943  TV-PG     61 min
8739      March 31, 2017         1943  TV-PG     82 min
...                  ...           ...    ...        ...
862         May 20, 2021         2021  TV-MA  2 Seasons
861         May 20, 2021         2021  TV-Y7     47 min
857         May 21, 2021         2021  TV-MA  2 Seasons
876         May 14, 2021         2021  TV-MA   1 Season
965       April 29, 2021         2021  TV-MA   1 Season

                              listed_in  \
4250                           TV Shows
7790          Classic Movies, Documentaries
8205          Classic Movies, Documentaries
8660          Classic Movies, Documentaries
8739                        Documentaries
...                                 ...
862                         TV Comedies
```

```
861                     Children & Family Movies, Comedies
857    International TV Shows, Spanish-Language TV Sh…
876                Docuseries, International TV Shows
965              Anime Series, International TV Shows

                                           description
4250   This collection restores films from women who …
7790   Frank Capra's documentary chronicles the rise …
8205   Director John Ford captures combat footage of …
8660   This World War II-era training film dramatizes…
8739   This installment of Frank Capra's acclaimed do…
…                                                    …
862    A young gay man with cerebral palsy branches o…
861    Little Singham and his crew travel to the year…
857    Self-centered Javier's life gets a bit messy w…
876    Alumni of India's elite academic institutes sh…
965    A peaceful boatman once known as the Black Sam…

[8807 rows x 12 columns]
```

[105]:
```python
# tail the df

net.tail(5)
```

[105]:
```
      show_id     type        title           director  \
8802   s8803    Movie        Zodiac    David Fincher
8803   s8804  TV Show   Zombie Dumb               NaN
8804   s8805    Movie    Zombieland  Ruben Fleischer
8805   s8806    Movie          Zoom     Peter Hewitt
8806   s8807    Movie        Zubaan      Mozez Singh

                                           cast        country  \
8802  Mark Ruffalo, Jake Gyllenhaal, Robert Downey J…  United States
8803                                             NaN            NaN
8804  Jesse Eisenberg, Woody Harrelson, Emma Stone, …  United States
8805  Tim Allen, Courteney Cox, Chevy Chase, Kate Ma…  United States
8806  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan…          India

              date_added  release_year rating   duration  \
8802   November 20, 2019          2007      R    158 min
8803        July 1, 2019          2018  TV-Y7  2 Seasons
8804    November 1, 2019          2009      R     88 min
8805    January 11, 2020          2006     PG     88 min
8806       March 2, 2019          2015  TV-14    111 min

                           listed_in  \
8802          Cult Movies, Dramas, Thrillers
```

```
8803           Kids' TV, Korean TV Shows, TV Comedies
8804                        Comedies, Horror Movies
8805          Children & Family Movies, Comedies
8806  Dramas, International Movies, Music & Musicals

                                         description
8802  A political cartoonist, a crime reporter and a…
8803  While living alone in a spooky town, a young g…
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…
```

0.0.2 Introduction explaining the data set

The dataset I choose is from Kaggle and is titled "Netflix Titles". The dataset is a comprehensive collection of movies and TV shows available on Netflix, with relaese dates between 1925 and 2021. The dataset has 8807 rows of data and 12 columns. The following is list of the columns.

show_id: A unique identifier for each title.

type: The category of the title, indicating whether it is a movie or a TV show.

title: The name of the movie or TV show.

director: The director(s) of the movie or TV show. This field may be blank for some TV shows where a single director is not attributed.

cast: A list of main actors/actresses involved in the title. This field may also be blank, indicating that the cast information is either not available or not applicable.

country: The country or countries where the movie or TV show was produced. It can contain multiple countries separated by commas.

date_added: The date on which the title was added to Netflix. This can vary across different countries but represents a general availability date.

release_year: The year in which the movie or TV show was originally released.

rating: The rating given to the title, indicating the appropriate audience age group based on content.

duration: The length of the movie in minutes or the number of seasons for TV shows.

listed_in: The genre(s) or category(ies) the title belongs to.

description: A brief description of the plot or premise of the title.

```
[106]: # the number NaN entries in each column

       nan_entries_per_column=net.isnull().sum()

       nan_entries_per_column  # display
```

```
[106]: 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
```

0.0.3 Question to explore

1. Content Growth Over Time: How has the number of titles (movies vs. TV shows) on Netflix changed over the years (looking at date added)

2. Release Patterns: Is there a trend in the release years of movies and TV shows.

3. Ratings Analysis: What ratings (e.g., PG, TV-MA) are most common among Netflix titles, and what does this suggest about the platform's target audience?

0.0.4 Create at least three graphs that help answer these questions.

Make sure your graphs are clearly readable and are labeled appropriately and professionally

```python
[107]: # Graph 1 fix data
       # convert 'date_added' to string type to ensure we can apply string methods
       net['date_added'] = pd.to_datetime(net['date_added'], errors='coerce')

       # extract the year from 'date_added' and create a new column 'year_added'
       net['year_added'] = net['date_added'].dt.year

       # fill na with 0
       net['year_added'] = net['year_added'].fillna(0).astype(int)
```

```python
[108]: # head changes

       net.head(15)
```

```
[108]:   show_id     type                          title  \
       0      s1    Movie           Dick Johnson Is Dead
       1      s2  TV Show                  Blood & Water
       2      s3  TV Show                      Ganglands
       3      s4  TV Show          Jailbirds New Orleans
       4      s5  TV Show                   Kota Factory
       5      s6  TV Show                  Midnight Mass
```

```
6      s7     Movie      My Little Pony: A New Generation
7      s8     Movie                               Sankofa
8      s9   TV Show       The Great British Baking Show
9     s10     Movie                          The Starling
10    s11   TV Show  Vendetta: Truth, Lies and The Mafia
11    s12   TV Show                       Bangkok Breaking
12    s13     Movie                          Je Suis Karl
13    s14     Movie      Confessions of an Invisible Girl
14    s15   TV Show        Crime Stories: India Detectives

                            director  \
0                     Kirsten Johnson
1                                 NaN
2                     Julien Leclercq
3                                 NaN
4                                 NaN
5                       Mike Flanagan
6     Robert Cullen, José Luis Ucha
7                       Haile Gerima
8                     Andy Devonshire
9                     Theodore Melfi
10                                NaN
11                  Kongkiat Komesiri
12                Christian Schwochow
13                      Bruno Garotti
14                                NaN


                                               cast  \
0                                                NaN
1   Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…
2   Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…
3                                                NaN
4   Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…
5   Kate Siegel, Zach Gilford, Hamish Linklater, H…
6   Vanessa Hudgens, Kimiko Glenn, James Marsden, …
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…
10                                               NaN
11  Sukollawat Kanarot, Sushar Manaying, Pavarit M…
12  Luna Wedler, Jannis Niewöhner, Milan Peschel, …
13  Klara Castanho, Lucca Picon, Júlia Gomes, Marc…
14                                               NaN

                                 country date_added  \
0                           United States 2021-09-25
1                            South Africa 2021-09-24
```

```
2                                          NaN 2021-09-24
3                                          NaN 2021-09-24
4                                        India 2021-09-24
5                                          NaN 2021-09-24
6                                          NaN 2021-09-24
7   United States, Ghana, Burkina Faso, United Kin… 2021-09-24
8                               United Kingdom 2021-09-24
9                                United States 2021-09-24
10                                         NaN 2021-09-24
11                                         NaN 2021-09-23
12                      Germany, Czech Republic 2021-09-23
13                                         NaN 2021-09-22
14                                         NaN 2021-09-22


    release_year rating   duration  \
0           2020  PG-13     90 min
1           2021  TV-MA  2 Seasons
2           2021  TV-MA   1 Season
3           2021  TV-MA   1 Season
4           2021  TV-MA  2 Seasons
5           2021  TV-MA   1 Season
6           2021     PG     91 min
7           1993  TV-MA    125 min
8           2021  TV-14  9 Seasons
9           2021  PG-13    104 min
10          2021  TV-MA   1 Season
11          2021  TV-MA   1 Season
12          2021  TV-MA    127 min
13          2021  TV-PG     91 min
14          2021  TV-MA   1 Season


                                          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 …
5                  TV Dramas, TV Horror, TV Mysteries
6                           Children & Family Movies
7     Dramas, Independent Movies, International Movies
8                        British TV Shows, Reality TV
9                                    Comedies, Dramas
10  Crime TV Shows, Docuseries, International TV S…
11  Crime TV Shows, International TV Shows, TV Act…
12                         Dramas, International Movies
13                 Children & Family Movies, Comedies
14      British TV Shows, Crime TV Shows, Docuseries
```

```
                                   description  year_added
0   As her father nears the end of his life, filmm…         2021
1   After crossing paths at a party, a Cape Town t…         2021
2   To protect his family from a powerful drug lor…         2021
3   Feuds, flirtations and toilet talk go down amo…         2021
4   In a city of coaching centers known to train I…         2021
5   The arrival of a charismatic young priest brin…         2021
6   Equestria's divided. But a bright-eyed hero be…         2021
7   On a photo shoot in Ghana, an American model s…         2021
8   A talented batch of amateur bakers face off in…         2021
9   A woman adjusting to life after a loss contend…         2021
10  Sicily boasts a bold "Anti-Mafia" coalition. B…         2021
11  Struggling to earn a living in Bangkok, a man …         2021
12  After most of her family is murdered in a terr…         2021
13  When the clever but socially-awkward Tetê join…         2021
14  Cameras following Bengaluru police on the job …         2021
```

[129]:
```python
# Graph 1

titles_by_year_type_cleaned = net[net['year_added'] > 0].groupby(['year_added',
 ↪'type']).size().unstack().fillna(0)


titles_by_year_type_cleaned.plot(kind='area')
plt.title('Number of Titles Added to Netflix Over the Years', fontsize=16)
plt.xlabel('Year Added', fontsize=16)
plt.ylabel('Number of Titles', fontsize=16)
plt.legend(title='Type')
plt.tight_layout()

plt.show()
```

Number of Titles Added to Netflix Over the Years
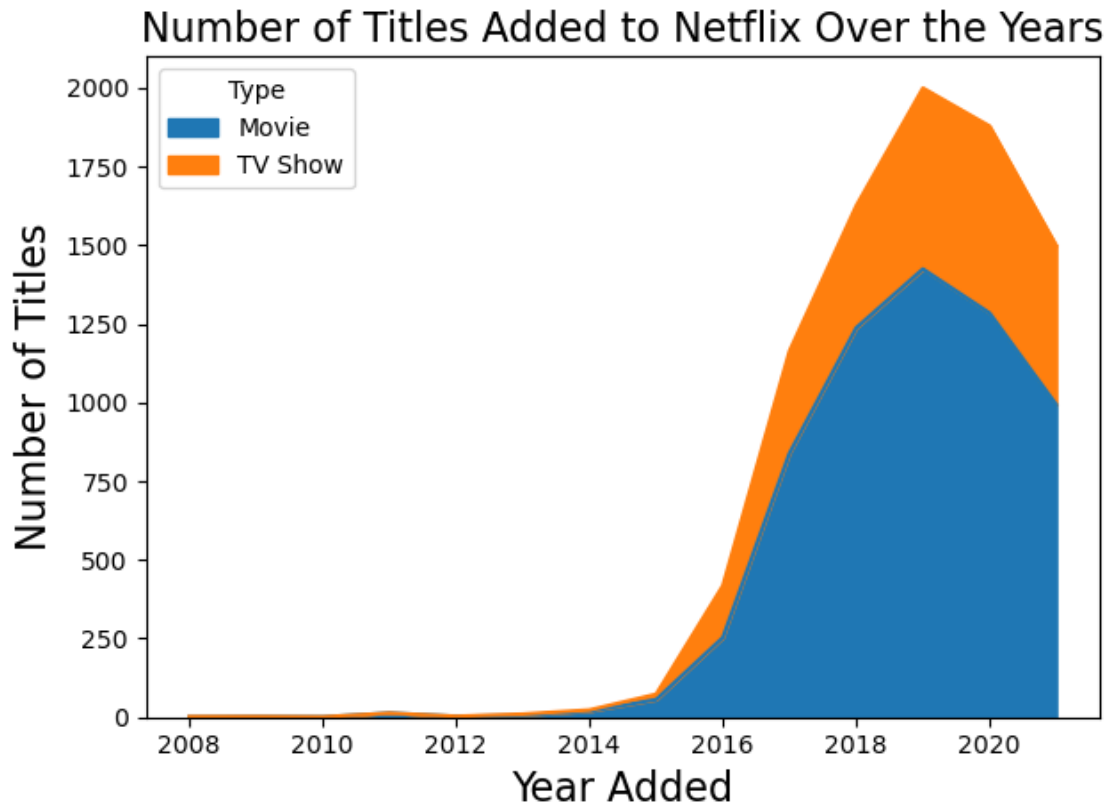
```
[110]:  # Graph 2
        # prepare the data

        release_counts = net[net['release_year'] > 0].groupby(['release_year', 'type']).
         ↪size().unstack(fill_value=0)
```

```
[111]:  # plot

        plt.figure(figsize=(12, 6))
        plt.plot(release_counts.index, release_counts['Movie'], label='Movies',⎵
         ↪marker='o', linestyle='-', color='blue')
        plt.plot(release_counts.index, release_counts['TV Show'], label='TV Shows',⎵
         ↪marker='x', linestyle='--', color='red')

        plt.title('Trend in Release Years of Netflix Movies and TV Shows', fontsize=16)
        plt.xlabel('Release Year', fontsize=14)
        plt.ylabel('Number of Titles', fontsize=14)
        plt.legend()
        plt.grid(True)
        plt.tight_layout()
```

```
plt.show()
```


Trend in Release Years of Netflix Movies and TV Shows
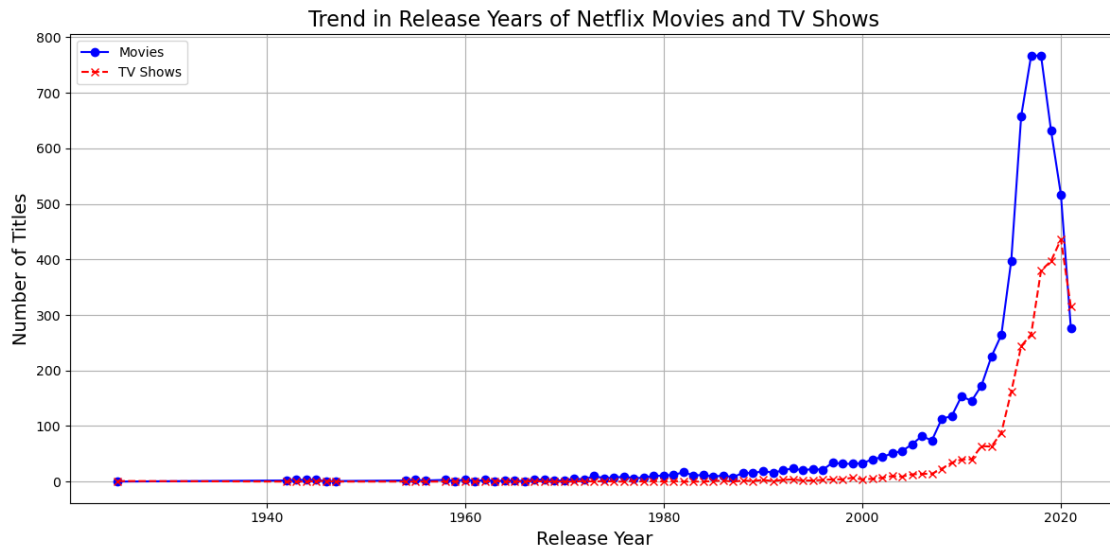
[117]:
```
# move 3 duration rows from rating to duration

# update the 'duration' column for each 'show_id'
net.loc[net['show_id'] == 's5542', 'duration'] = '74 min'
net.loc[net['show_id'] == 's5795', 'duration'] = '84 min'
net.loc[net['show_id'] == 's5814', 'duration'] = '66 min'

# update the 'duration' column for each 'show_id'
net.loc[net['show_id'] == 's5542', 'rating'] = 'na'
net.loc[net['show_id'] == 's5795', 'rating'] = 'na'
net.loc[net['show_id'] == 's5814', 'rating'] = 'na'
```

[119]:
```
# check changes print row

print(net[net['show_id'] == 's5542'])
```

```
      show_id    type            title    director         cast          country  \
5541   s5542   Movie  Louis C.K. 2017  Louis C.K.  Louis C.K.  United States

      date_added  release_year rating duration listed_in  \
5541  2017-04-04          2017     na   74 min    Movies

                                     description  year_added
5541  Louis C.K. muses on religion, eternal love, gi…        2017
```

```
[134]:  # Graph 3
        # prepare the data

        # drop the null/na
        df_clean = net.dropna(subset=['rating'])

        ratings_count = df_clean['rating'].value_counts()

        sorted_ratings_count = ratings_count.sort_values(ascending=False)
```
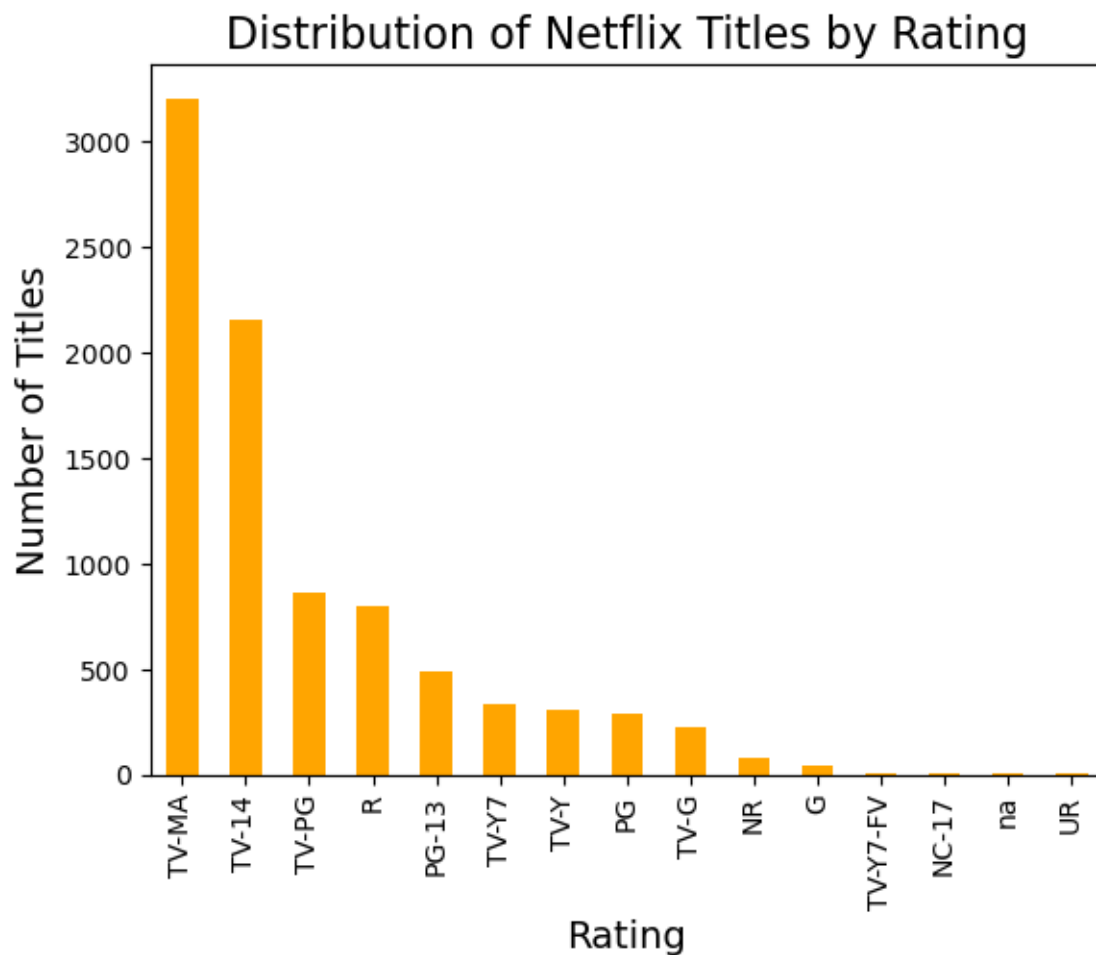
```
[135]:  # Plot

        sorted_ratings_count.plot(kind='bar', color='orange')

        plt.title('Distribution of Netflix Titles by Rating', fontsize=16)
        plt.xlabel('Rating', fontsize=14)
        plt.ylabel('Number of Titles', fontsize=14)

        plt.show()
```


Distribution of Netflix Titles by Rating

0.0.5 Explain what you have learned from each of your graphs.

Graph 1 is a area graph and compares the number of titles added to Netflix from 2008 to 2021. The graph shows year of year growth between 2013 and 2019 for Movies and 2013 to 2020 for TV Shows, with movies growing at a more than double clip between 2017 to 2021. There is slight downturn for movies from 2020 to 2021 and only a slight down turn for TV Shows for 2021.

Graph 2 is a line graph with markers. It compares the number of Netflix movies and TV shows released over time, with the release years on the x-axis and the number of titles on the y-axis. Movies are represented by a blue line with circle markers and a solid line style, while TV shows are indicated by a red line with 'x' markers and a dashed line style. The graph shows that early released years Movies and TV Shows trended equally together. However, begining with release years in 1980 Movies grew at a higher clip than TV Showes with this trend maintaining through the late 2000 teens, with both Movies and TV Shows dropping dramatically by 2020.

Graph 3 is an bar graph that displays the distributon of Netflix titles by rating. By far the most popular rating of Movies and TV Shows is TV-MA, followed by TV-14, TV-PG, R, PG-13,TV-Y7, TV-Y, PG, TV_G. The other ratings were negligable.

0.0.6 Conclusion

Content Growth Over Time

The analysis of how the number of titles on Netflix has changed over the years shows a significant increase in both movies and TV shows, with perhaps a more pronounced growth in recent years (excluding the COVID time). This could be attributed to Netflix's shift towards becoming a primary content creator and distributor, leading to an aggressive expansion of its library to attract and retain subscribers. The difference in growth rates between movies and TV shows reflects strategic decisions by Netflix to invest more heavily in one type of content based on viewer preferences and market trends.

Release Patterns

The trend in the release years of movies and TV shows on Netflix likely indicates an increasing focus on recent content, although a substantial number of older titles remain important for catalog diversity. There might be spikes in certain years, correlating with Netflix's investments in original content or exclusive streaming rights acquisitions. The analysis could reveal that while Netflix places a strong emphasis on new releases to keep its library fresh and engaging, it also values classic and evergreen titles that cater to a broad audience spectrum.

Ratings Analysis

The most common ratings among Netflix titles, such as PG, TV-MA, etc., suggest a targeted approach by Netflix towards certain demographic groups. For instance, the predominance of TV-MA (Mature Audience) rated titles indicates Netflix's focus on adult viewers, offering a wide range of content from intense dramas to adult comedies and documentaries. The lower ratings like PG or PG-13 reflects Netflix's strategy to not overly cater to families and younger viewers with a variety of movies and shows that are suitable for a younger audience. The diversity in ratings can highlight

Netflix's intention to appeal to a global audience with varied tastes and preferences, aiming to be a one-stop entertainment solution for people of all ages.

Overall Conclusion

The analysis based on the graphs produced in the notebook reveals Netflix's evolving content strategy, emphasizing a balanced mix of movies and TV shows to cater to a wide audience. The growth in titles over the years underscores Netflix's commitment to expanding its library, both through acquisitions and the production of original content. Release patterns hint at a strategic focus on fresh and relevant content while maintaining a selection of classics and evergreen titles. Finally, the ratings analysis suggests that Netflix aims to serve a more of an adult audience base.

[ ]: