task6

July 4, 2024

```python
[ ]: '''1. Implement time series analysis for forecasting trends and seasonality.
2. Perform sentiment analysis or text mining on unstructured data.
3. Explore clustering or classification techniques for segmentation and pattern
recognition.'''
```

```python
[2]: %pip install seaborn
```

```python
[3]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```python
[4]: disney_data = pd.read_csv("disney_plus_titles.csv")
```

```python
[5]: type(disney_data)
```

```
[5]: pandas.core.frame.DataFrame
```

```python
[6]: disney_data.head()
```

```
[6]:   show_id      type                         title  \
    0      s1     Movie                 A Spark Story
    1      s2     Movie                 Spooky Buddies
    2      s3     Movie           The Fault in Our Stars
    3      s4   TV Show                 Dog: Impossible
    4      s5   TV Show   Spidey And His Amazing Friends

                      director  \
    0  Jason Sterman, Leanne Dare
    1               Robert Vince
    2                 Josh Boone
    3                        NaN
    4                        NaN

                                                 cast                  country  \
    0                     Apthon Corbin, Louis Gonzales                     NaN
    1  Tucker Albrizzi, Diedrich Bader, Ameko Eks Mas…  United States, Canada
    2  Shailene Woodley, Ansel Elgort, Laura Dern, Sa…          United States
```

```
3                                          Matt Beisner           United States
4   Benjamin Valic, Lily Sanfelippo, Jakari Fraser…           United States


            date_added  release_year rating   duration  \
0  September 24, 2021          2021  TV-PG      88 min
1  September 24, 2021          2011      G      93 min
2  September 24, 2021          2014  PG-13     127 min
3  September 22, 2021          2019  TV-PG   2 Seasons
4  September 22, 2021          2021   TV-Y    1 Season


                            listed_in  \
0                         Documentary
1                 Comedy, Fantasy, Kids
2          Coming of Age, Drama, Romance
3  Animals & Nature, Docuseries, Family
4     Action-Adventure, Animation, Kids


                                  description
0  Two Pixar filmmakers strive to bring their uni…
1  The puppies go on a spooky adventure through a…
2  Hazel and Gus share a love that sweeps them on…
3  Matt Beisner uses unique approaches to modifyi…
4  Spidey teams up with pals to become The Spidey…
```

[8]: `disney_data.info`

```
[8]: <bound method DataFrame.info of       show_id      type
     title  \
0          s1    Movie                   A Spark Story
1          s2    Movie                   Spooky Buddies
2          s3    Movie             The Fault in Our Stars
3          s4  TV Show                   Dog: Impossible
4          s5  TV Show  Spidey And His Amazing Friends
…           …       …                               …
1363   s1364    Movie             The Sword in the Stone
1364   s1365    Movie                   Those Calloways
1365   s1366  TV Show             Disney Kirby Buckets
1366   s1367  TV Show                   Disney Mech-X4
1367   s1368  TV Show             Imagination Movers


                         director  \
0      Jason Sterman, Leanne Dare
1                   Robert Vince
2                     Josh Boone
3                            NaN
4                            NaN
…                              …
```

```
1363           Wolfgang Reitherman
1364               Norman Tokar
1365                       NaN
1366                       NaN
1367                       NaN

                                                   cast  \
0                         Apthon Corbin, Louis Gonzales
1     Tucker Albrizzi, Diedrich Bader, Ameko Eks Mas…
2     Shailene Woodley, Ansel Elgort, Laura Dern, Sa…
3                                         Matt Beisner
4     Benjamin Valic, Lily Sanfelippo, Jakari Fraser…
…                                                   …
1363  Sebastian Cabot, Karl Swenson, Rickie Sorensen…
1364  Brian Keith, Vera Miles, Brandon de Wilde, Wal…
1365  Jacob Bertrand, Mekai Curtis, Cade Sutton, Oli…
1366  Nathaniel Potvin, Raymond Cham, Kamran Lucas, …
1367  Rich Collins, Dave Poche, Scott Durbin, Scott …

                    country          date_added  release_year rating  \
0                       NaN  September 24, 2021          2021  TV-PG
1     United States, Canada  September 24, 2021          2011      G
2             United States  September 24, 2021          2014  PG-13
3             United States  September 22, 2021          2019  TV-PG
4             United States  September 22, 2021          2021   TV-Y
…                         …                   …             …      …
1363          United States     October 1, 2019          1963      G
1364          United States     October 1, 2019          1965     PG
1365          United States                 NaN          2014  TV-Y7
1366                 Canada                 NaN          2016  TV-Y7
1367          United States                 NaN          2008   TV-Y

        duration                                  listed_in  \
0         88 min                                Documentary
1         93 min                      Comedy, Fantasy, Kids
2        127 min              Coming of Age, Drama, Romance
3      2 Seasons         Animals & Nature, Docuseries, Family
4       1 Season          Action-Adventure, Animation, Kids
…              …                                          …
1363      80 min        Action-Adventure, Animation, Comedy
1364     132 min              Animals & Nature, Drama, Family
1365   3 Seasons    Action-Adventure, Comedy, Coming of Age
1366   2 Seasons  Action-Adventure, Comedy, Science Fiction
1367   3 Seasons                                Kids, Music

                                            description
0     Two Pixar filmmakers strive to bring their uni…
```

```
1     The puppies go on a spooky adventure through a…
2     Hazel and Gus share a love that sweeps them on…
3     Matt Beisner uses unique approaches to modifyi…
4     Spidey teams up with pals to become The Spidey…
…                                                 …
1363  Merlin trains a young orphan who's destined to…
1364  A strong-willed family struggles to establish …
1365   Welcome to Kirby's world! It's rude and sketchy.
1366  Ryan discovers his ability to control a giant …
1367  Rock out with the Imagination Movers, Disney J…

[1368 rows x 12 columns]>
```

[9]: `disney_data.describe()`

```
[9]:        release_year
      count   1368.000000
      mean    2002.348684
      std       22.127559
      min     1928.000000
      25%     1998.000000
      50%     2011.000000
      75%     2018.000000
      max     2021.000000
```

[11]: `disney_data.isnull()`

```
[11]:        show_id   type   title  director   cast  country  date_added  \
      0       False  False  False     False  False     True       False
      1       False  False  False     False  False    False       False
      2       False  False  False     False  False    False       False
      3       False  False  False      True  False    False       False
      4       False  False  False      True  False    False       False
      …          …      …      …         …      …        …
      1363    False  False  False     False  False    False       False
      1364    False  False  False     False  False    False       False
      1365    False  False  False      True  False    False        True
      1366    False  False  False      True  False    False        True
      1367    False  False  False      True  False    False        True

            release_year  rating  duration  listed_in  description
      0            False   False     False      False        False
      1            False   False     False      False        False
      2            False   False     False      False        False
      3            False   False     False      False        False
      4            False   False     False      False        False
      …               …      …         …          …            …
```

```
1363          False   False     False       False         False
1364          False   False     False       False         False
1365          False   False     False       False         False
1366          False   False     False       False         False
1367          False   False     False       False         False

[1368 rows x 12 columns]
```
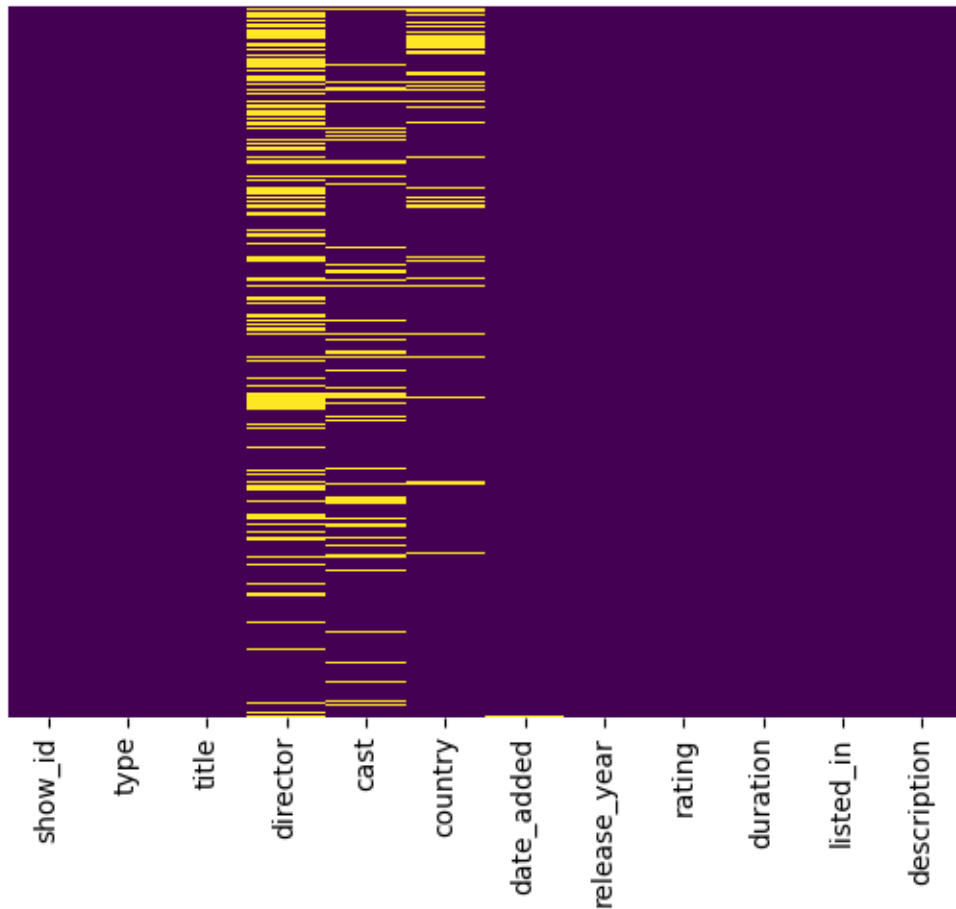
[12]: `disney_data.isnull().sum()`

```
[12]: show_id          0
      type             0
      title            0
      director       440
      cast           174
      country        175
      date_added       3
      release_year     0
      rating           2
      duration         0
      listed_in        0
      description      0
      dtype: int64
```

[13]: `sns.heatmap(disney_data.isnull(),yticklabels=False,cbar=False,cmap='viridis')`

[13]: `<AxesSubplot:>`

```
[14]: sns.set_style("whitegrid")
```

```
[15]: genre_count = disney_data['listed_in'].str.split(', ').explode().value_counts()
      plt.figure(figsize=(12, 6))
      sns.barplot(x=genre_count.head(10).index, y=genre_count.head(10).
        ↪values,hue=genre_count.head(10).index, palette="viridis", dodge=False)
      plt.legend([],[], frameon=False)
      plt.title('Top 10 Most Frequent Genres on Disney Plus', fontsize=16)
      plt.xlabel('Genre', fontsize=14)
      plt.ylabel('Frequency', fontsize=14)
      plt.xticks(rotation=45, fontsize=12)
      plt.yticks(fontsize=12)
      plt.show()
```

Top 10 Most Frequent Genres on Disney Plus
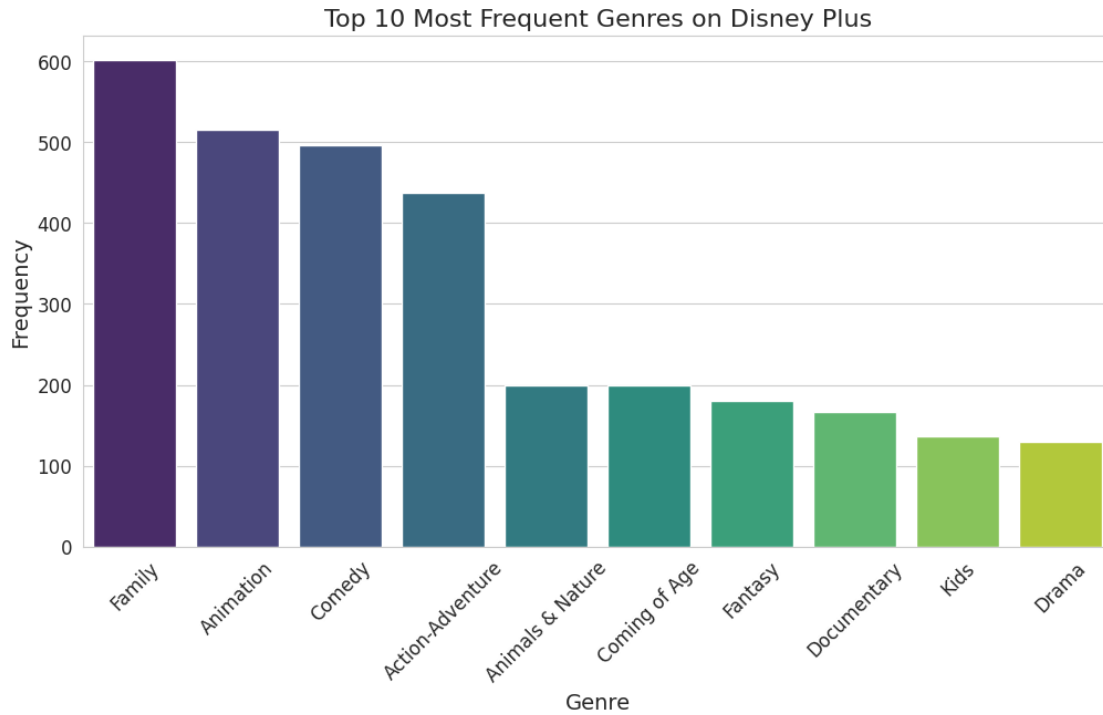
```
[16]: disney_data['date_added'] = pd.to_datetime(disney_data['date_added'])
      rating_trend = disney_data.groupby([disney_data['date_added'].dt.
       ↪year,'rating']).size().unstack().fillna(0)
      plt.figure(figsize=(12, 6))
      sns.lineplot(data=rating_trend)
      plt.title('Distribution of Content Ratings Over Time', fontsize=16)
      plt.xlabel('Year', fontsize=14)
      plt.ylabel('Number of Titles', fontsize=14)
      plt.xticks(fontsize=12)
      plt.yticks(fontsize=12)
      plt.legend(title='Rating', fontsize=12, title_fontsize=14)
      plt.show()
```

Distribution of Content Ratings Over Time
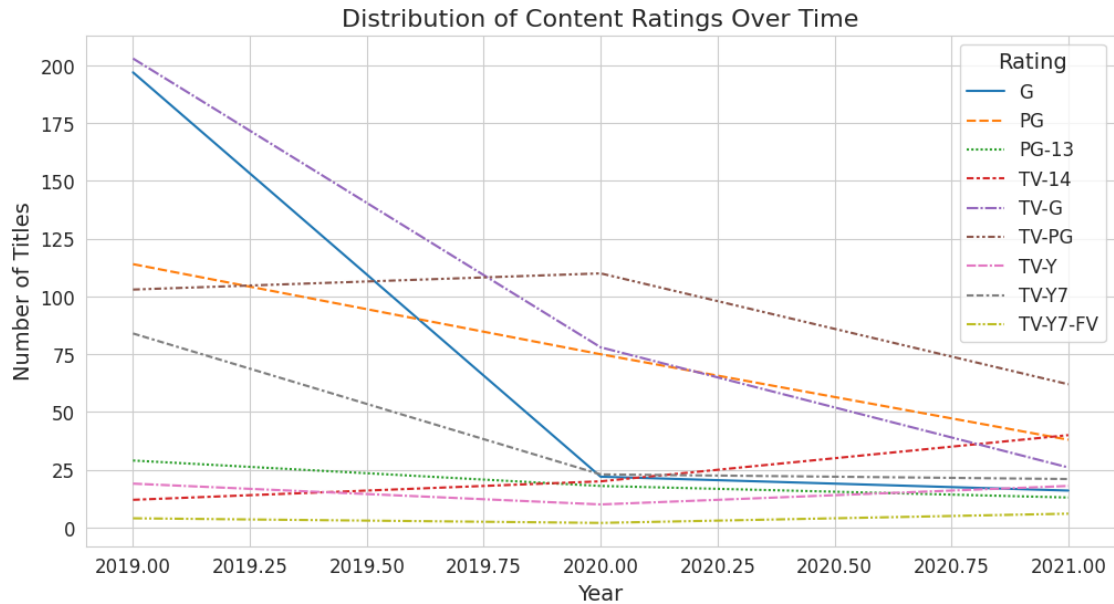
```
country_genre_count = disney_data['country'].str.split(', ').explode().
 ↪value_counts()
plt.figure(figsize=(12, 6))
sns.barplot(x=country_genre_count.head(10).index, y=country_genre_count.
 ↪head(10).values, hue=country_genre_count.head(10).index,␣
 ↪palette="magma",dodge=False)
plt.legend([],[], frameon=False)
plt.title('Top 10 Countries Producing Content on Disney Plus', fontsize=16)
plt.xlabel('Country', fontsize=14)
plt.ylabel('Number of Titles', fontsize=14)
plt.xticks(rotation=45, fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```

Top 10 Countries Producing Content on Disney Plus
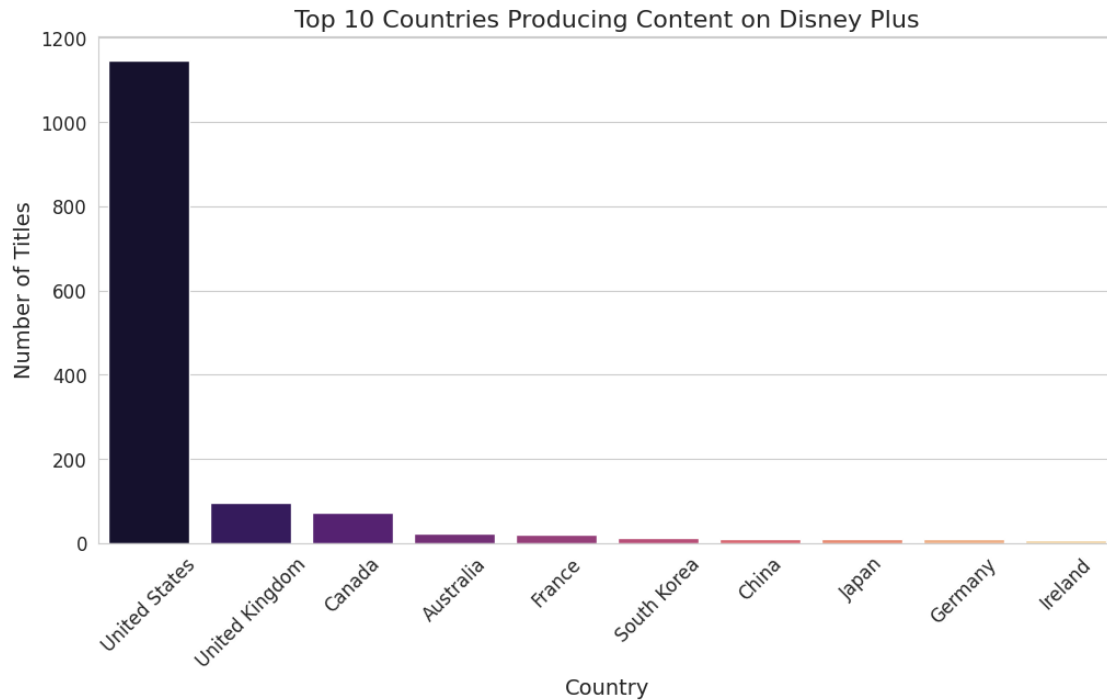
```
[18]: titles_per_year = disney_data['date_added'].dt.year.value_counts().sort_index()
      plt.figure(figsize=(12, 6))
      sns.barplot(x=titles_per_year.index, y=titles_per_year.
        ↪values,hue=titles_per_year.index, palette="coolwarm", dodge=False)
      plt.legend([],[], frameon=False)
      plt.title('Number of New Titles Added Per Year', fontsize=16)
      plt.xlabel('Year', fontsize=14)
      plt.ylabel('Number of Titles', fontsize=14)
      plt.xticks(rotation=45, fontsize=12)
      plt.yticks(fontsize=12)
      plt.show()
```

9

Number of New Titles Added Per Year
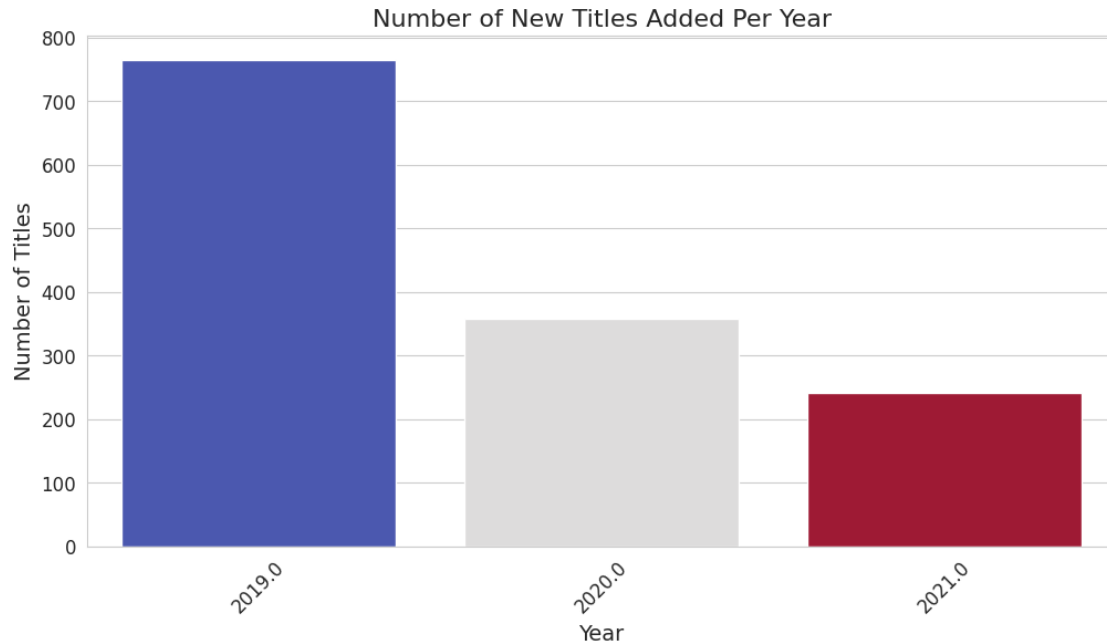
```
[19]: director_count = disney_data['director'].value_counts().head(10)
      plt.figure(figsize=(12, 6))
      sns.barplot(x=director_count.index, y=director_count.values, hue=director_count.
       ↪index, palette="cubehelix", dodge=False)
      plt.legend([],[], frameon=False)
      plt.title('Top 10 Most Prolific Directors on Disney Plus', fontsize=16)
      plt.xlabel('Director', fontsize=14)
      plt.ylabel('Number of Titles', fontsize=14)
      plt.xticks(rotation=45, fontsize=12)
      plt.yticks(fontsize=12)
      plt.show()
```

Top 10 Most Prolific Directors on Disney Plus
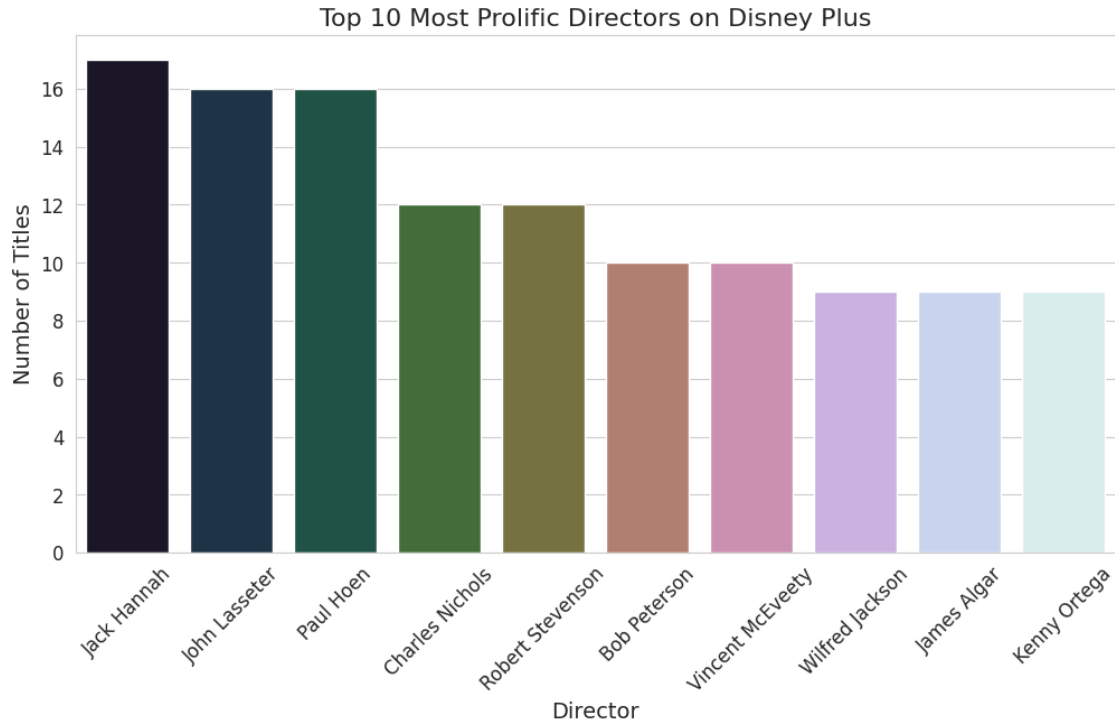

```
[20]: director_genre = disney_data.explode('listed_in').
     ↪groupby('director')['listed_in'].apply(lambda x: ', '.join(x.unique()))
     print("Directors and Their Associated Genres:\n", director_genre.head(10))
```

Directors and Their Associated Genres:
 director
Aaron Blaise, Robert Walker   Action-Adventure, Animation,
Family
Adam Shankman   Action-Adventure, Comedy, Spy/Espionage,
Comed…
Adam Stein, Zach Lipovsky   Action-Adventure, Comedy, Coming
of Age
Alan Barillaro   Animals & Nature, Animation,
Family
Alan Shapiro   Crime, Drama,
Family
Alan Taylor   Action-Adventure, Fantasy, Science
Fiction
Alastair Fothergill, Andy Byatt   Animals & Nature, Documentary,
Family
Alastair Fothergill, Jeff Wilson   Animals & Nature, Documentary,
Family
Alastair Fothergill, Keith Scholey   Animals & Nature, Documentary,
Family

```
Alastair Fothergill, Mark Linfield          Animals & Nature, Documentary,
Family
Name: listed_in, dtype: object
```

[22]:
```python
disney_data['duration_minutes'] = disney_data['duration'].apply(lambda x: int(x.
 ↪split()[0]) if 'min' in x else None)
average_duration = disney_data.groupby('type')['duration_minutes'].mean()
plt.figure(figsize=(12, 6))
sns.barplot(x=average_duration.index, y=average_duration.
 ↪values,hue=average_duration.index, palette="pastel", dodge=False)
plt.legend([],[], frameon=False)
plt.title('Average Duration of Movies vs. TV Shows on Disney Plus', fontsize=16)
plt.xlabel('Type', fontsize=14)
plt.ylabel('Average Duration (minutes)', fontsize=14)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```
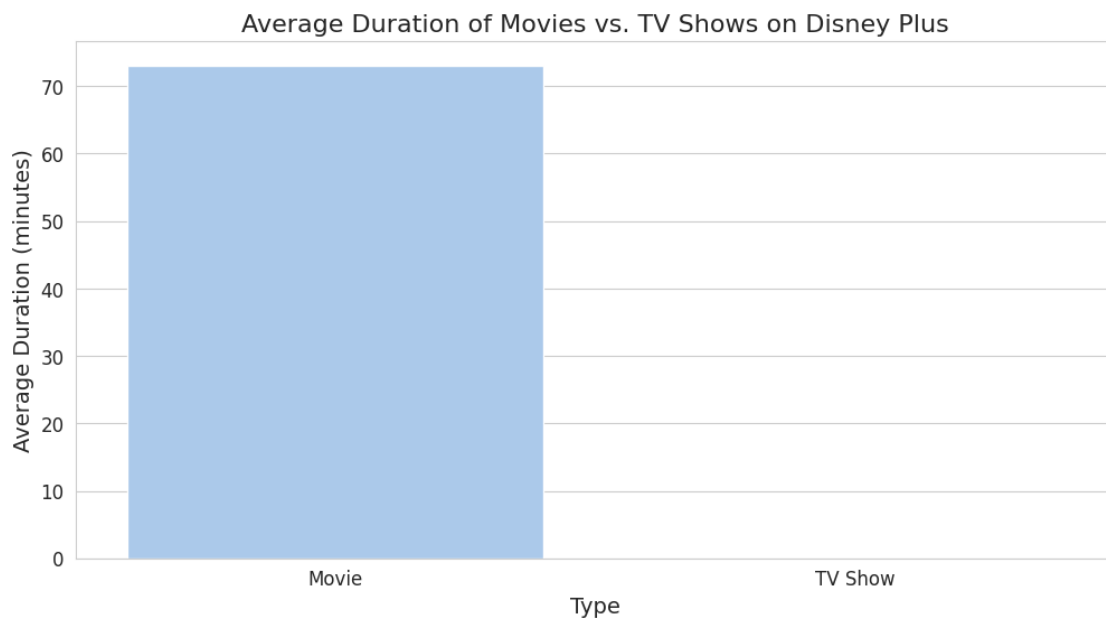

[23]:
```python
disney_data['cast_size'] = disney_data['cast'].apply(lambda x: len(x.
 ↪split(',')) if pd.notnull(x) else 0)
cast_size_distribution = disney_data.groupby('type')['cast_size'].mean()
plt.figure(figsize=(12, 6))
sns.barplot(x=cast_size_distribution.index, y=cast_size_distribution.
 ↪values,hue=cast_size_distribution.index, palette="muted", dodge=False)
plt.legend([],[], frameon=False)
plt.title('Average Cast Size of Movies vs. TV Shows on Disney Plus',fontsize=16)
plt.xlabel('Type', fontsize=14)
```

```
plt.ylabel('Average Cast Size', fontsize=14)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```
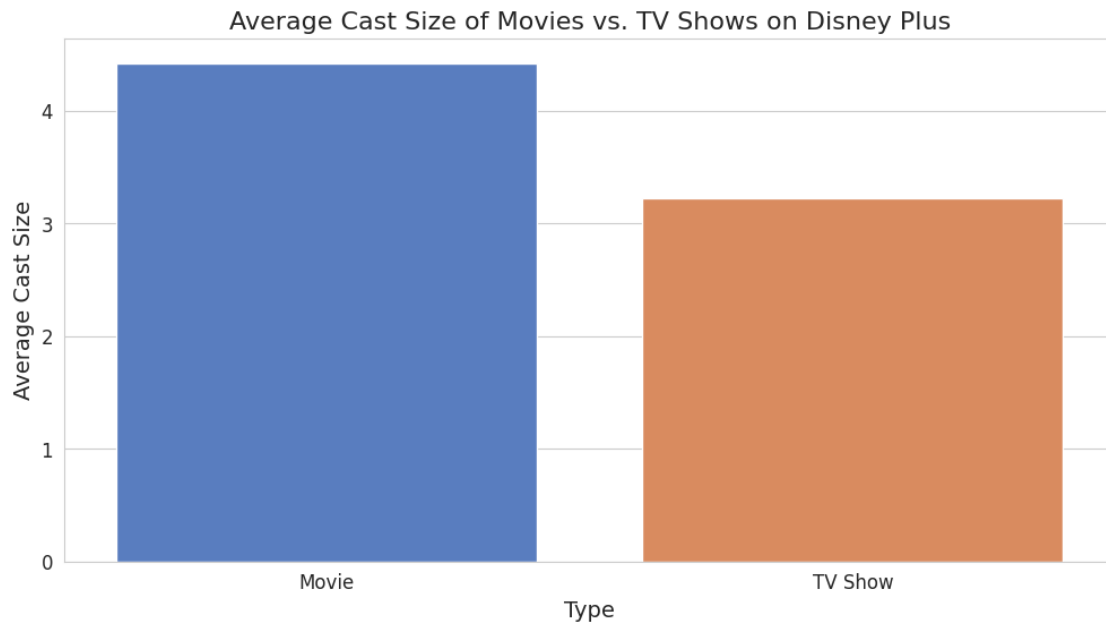

[24]:
```
plt.figure(figsize=(12, 6))
sns.histplot(disney_data[disney_data['type'] == 'Movie']['duration_minutes'].
 ↪dropna(), bins=30, kde=True, color='dodgerblue')
plt.title('Distribution of Movie Durations', fontsize=16)
plt.xlabel('Duration (minutes)', fontsize=14)
plt.ylabel('Frequency', fontsize=14)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```

Distribution of Movie Durations
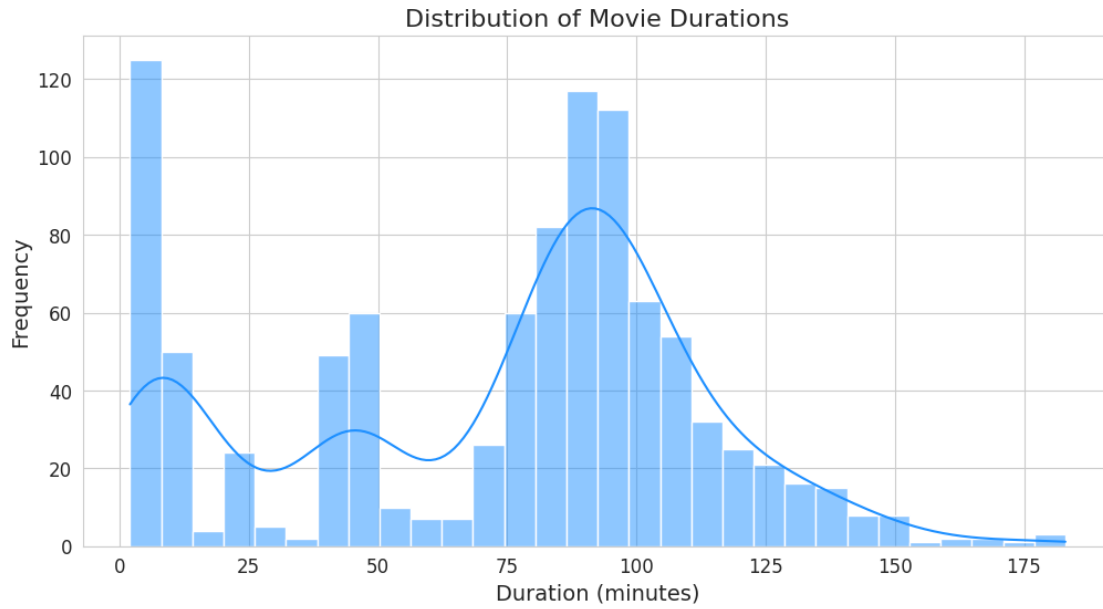

```
[25]: plt.figure(figsize=(12, 6))
      sns.boxplot(x='type', y='duration_minutes', data=disney_data)
      plt.title('Boxplot of Duration by Content Type', fontsize=16)
      plt.xlabel('Content Type', fontsize=14)
      plt.ylabel('Duration (minutes)', fontsize=14)
      plt.xticks(fontsize=12)
      plt.yticks(fontsize=12)
      plt.show()
```

Boxplot of Duration by Content Type
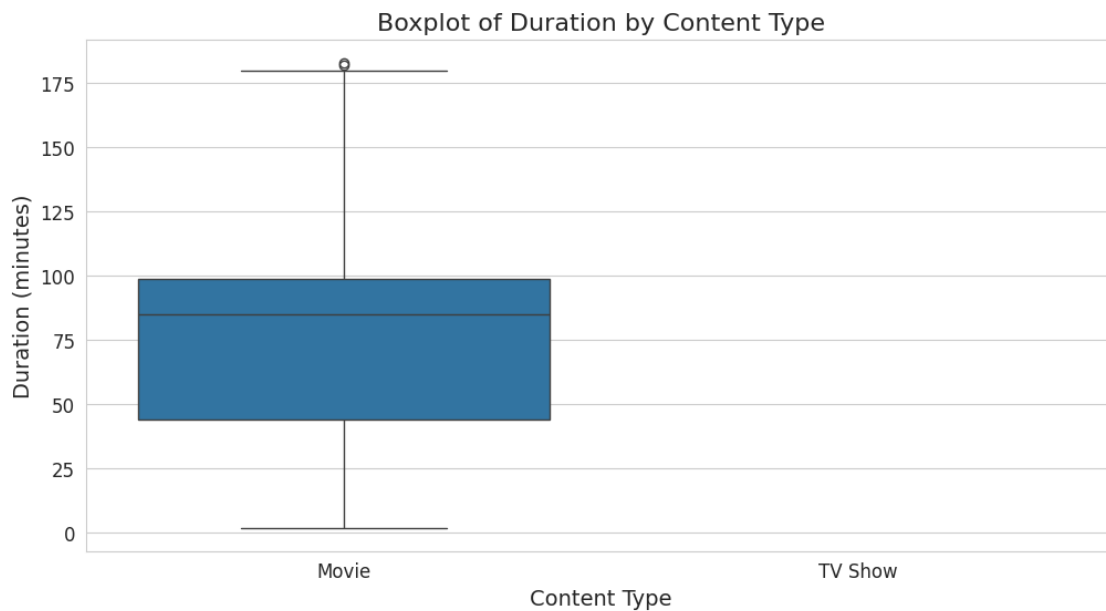
```
[26]: plt.figure(figsize=(14, 10))
      corr_matrix = disney_data[['duration_minutes', 'cast_size']].corr()
      sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)
      plt.title('Correlation Matrix', fontsize=16)
      plt.xticks(fontsize=12, rotation=45)
      plt.yticks(fontsize=12)
      plt.show()
```
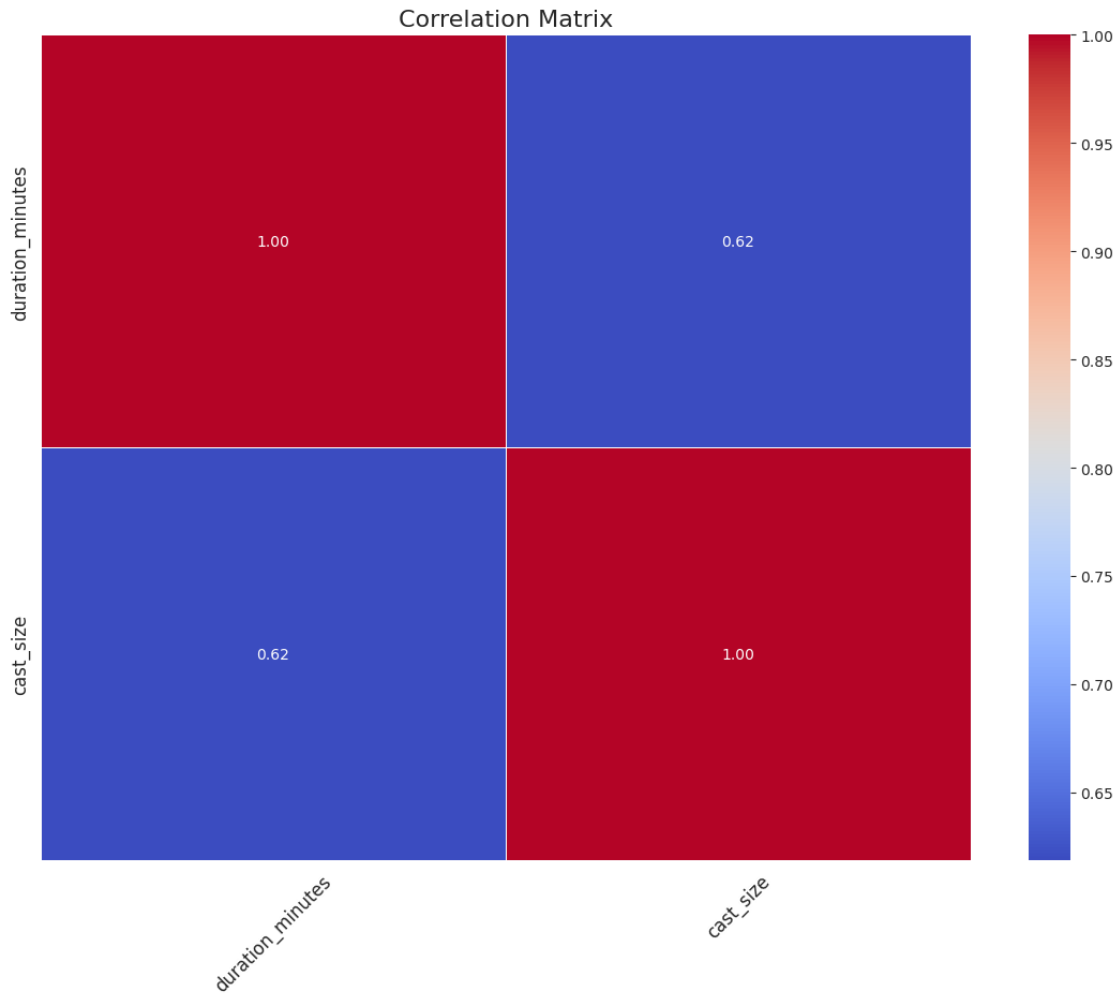


```
[27]: plt.xlabel('Release Year')
      plt.ylabel('Count')
      plt.title('Distribution of Release Years')
```
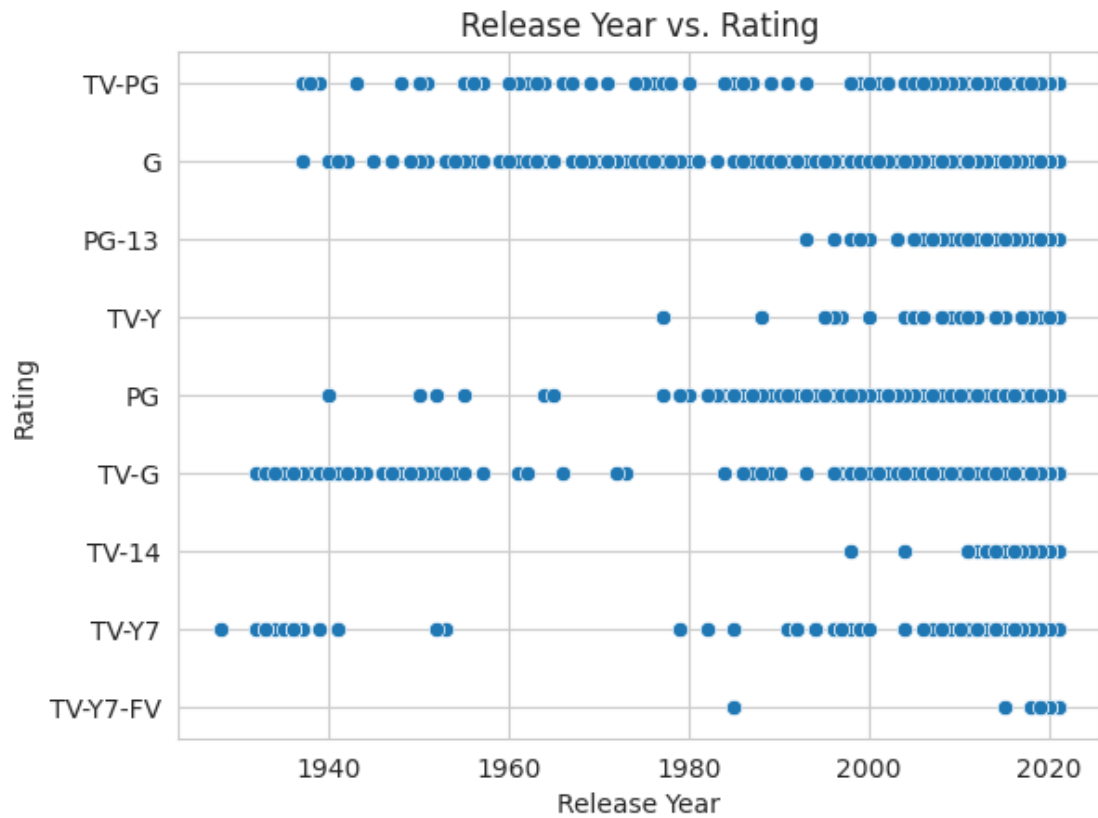
[27]: Text(0.5, 1.0, 'Distribution of Release Years')

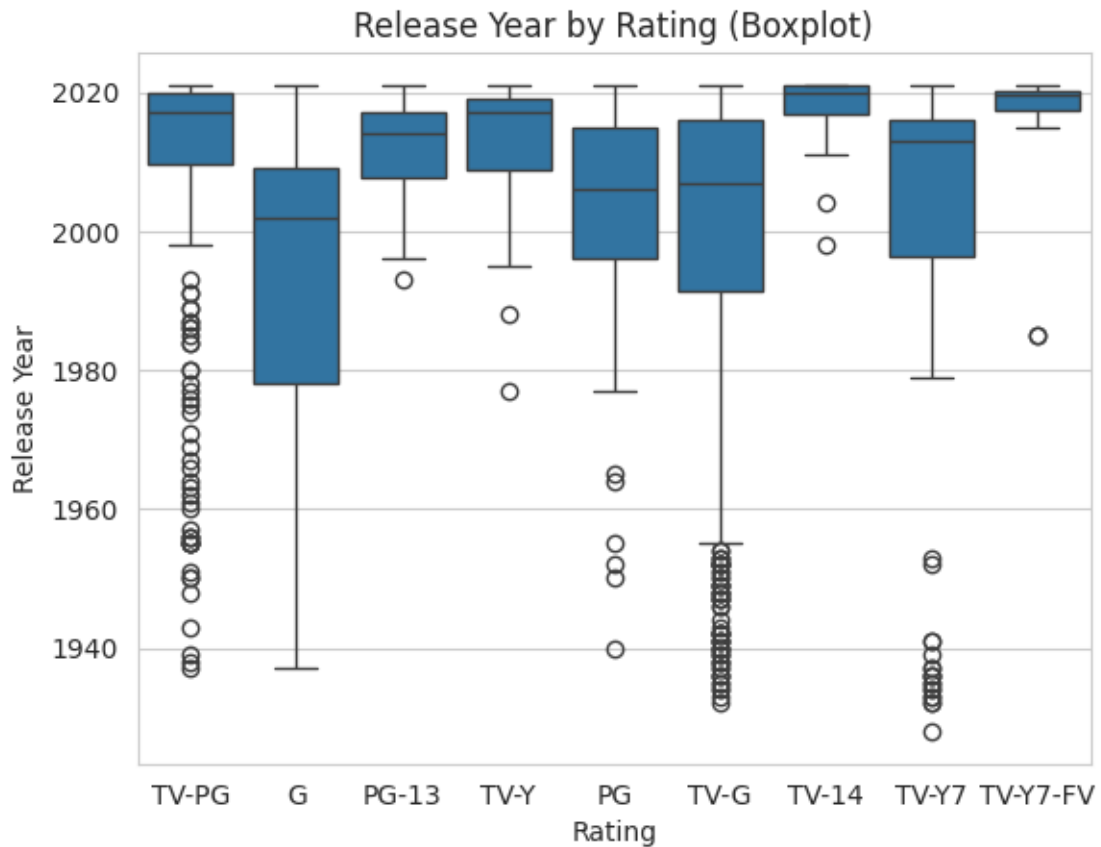## Distribution of Release Years
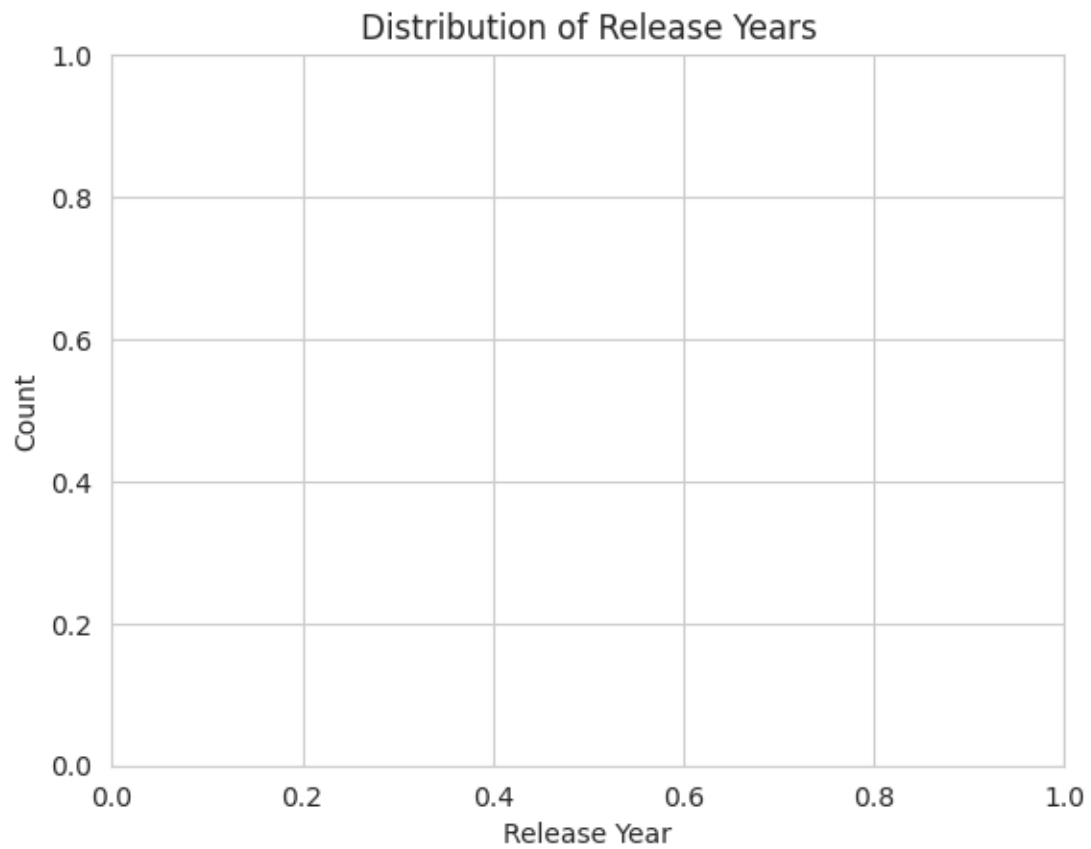

```
[28]:  disney_data['rating'].value_counts().plot(kind='bar')
       plt.xlabel('Rating')
       plt.ylabel('Count')
       plt.title('Distribution of Ratings')
       plt.show()
```
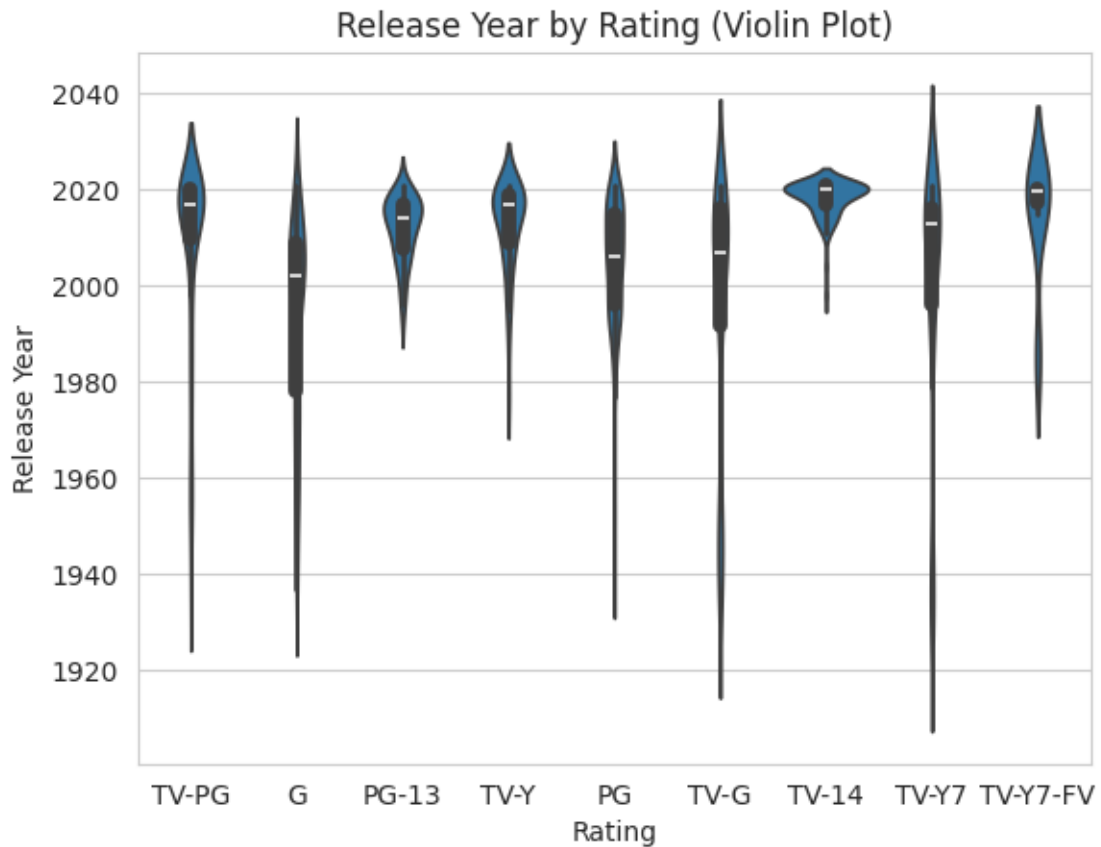
Distribution of Ratings
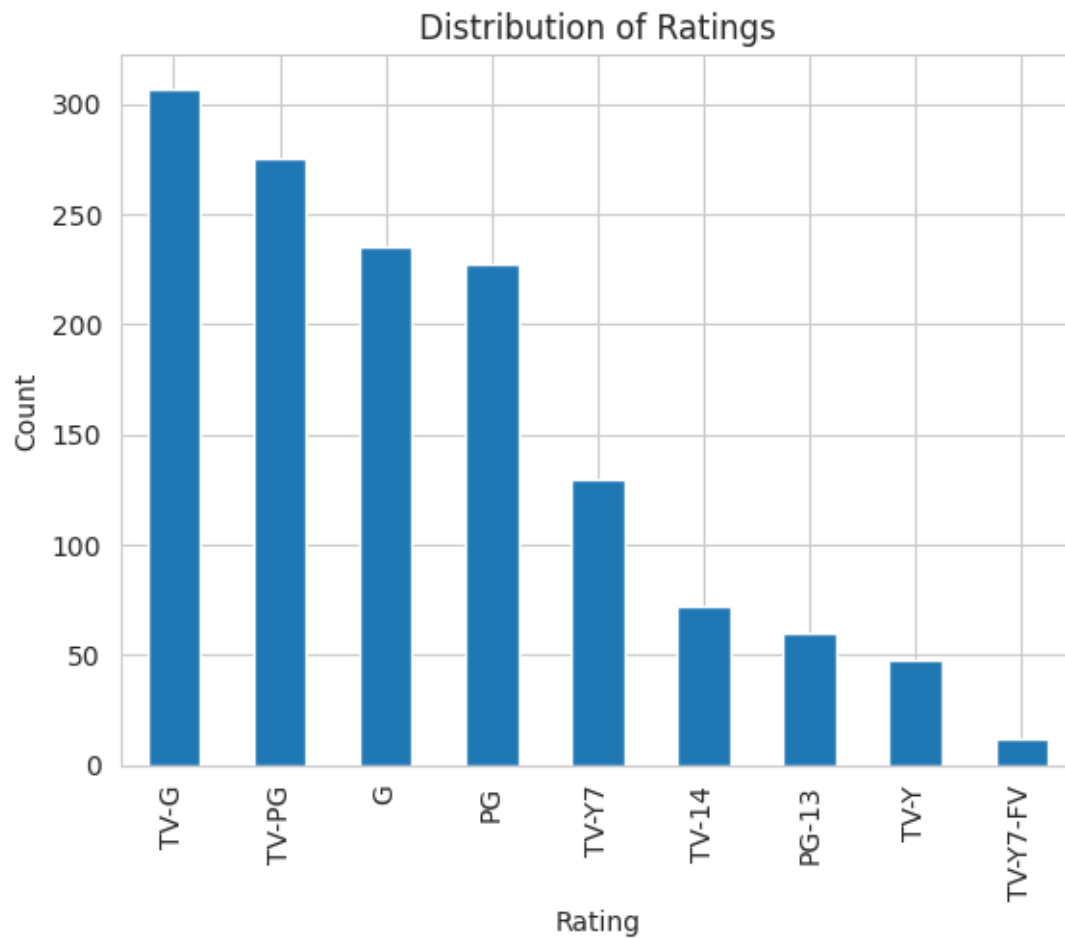

```
[29]: sns.scatterplot(data=disney_data, x='release_year', y='rating')
      plt.xlabel('Release Year')
      plt.ylabel('Rating')
      plt.title('Release Year vs. Rating')
      plt.show()
```

Release Year vs. Rating

```
sns.boxplot(x='rating', y='release_year', data=disney_data)
plt.xlabel('Rating')
plt.ylabel('Release Year')
plt.title('Release Year by Rating (Boxplot)')
plt.show()
```

Release Year by Rating (Boxplot)


```
[31]: sns.violinplot(x='rating', y='release_year', data=disney_data)
      plt.xlabel('Rating')
      plt.ylabel('Release Year')
      plt.title('Release Year by Rating (Violin Plot)')
      plt.show()
```

Release Year by Rating (Violin Plot)


```
sns.histplot(data=disney_data, x='release_year', bins=10)
plt.xlabel('Release Year')
plt.ylabel('Count')
plt.title('Distribution of Release Years (Histogram)')
plt.show()
```

## Distribution of Release Years (Histogram)
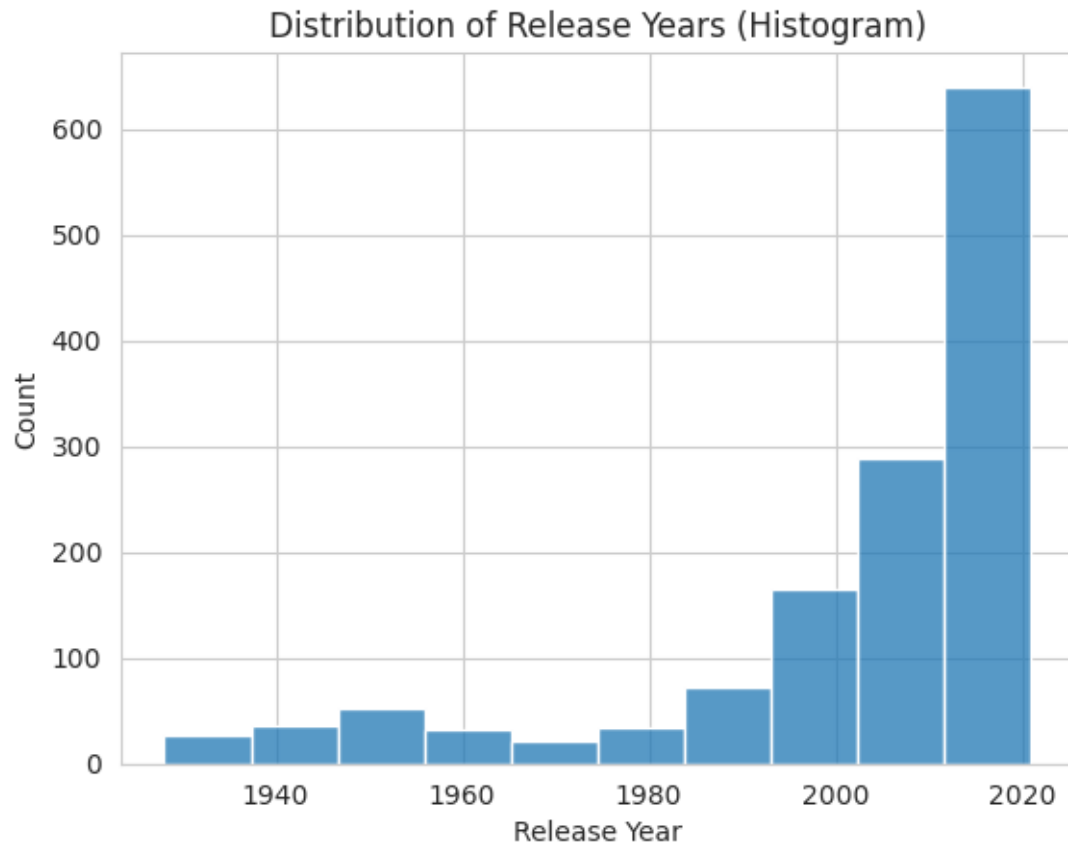

```
[33]: labels = disney_data['rating'].value_counts().index.to_list()
      sizes = disney_data['rating'].value_counts().to_list()
      colors = sns.color_palette('husl', len(labels))
      plt.pie(sizes, labels=labels, colors=colors, autopct='%.1f%%')
      plt.title('Distribution of Ratings (Pie Chart)')
      plt.show()
```

Distribution of Ratings (Pie Chart)
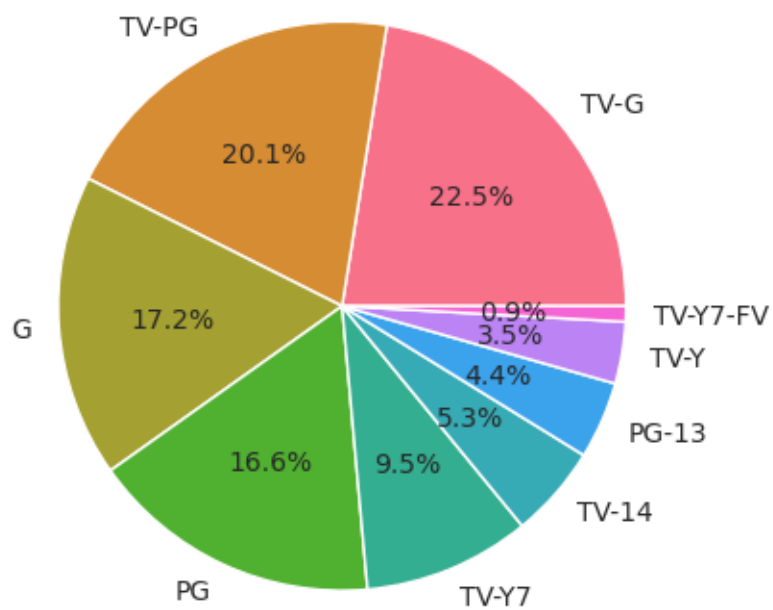

```
[34]: sns.scatterplot(data=disney_data, x='release_year', y='rating')
      plt.xlabel('Release Year')
      plt.ylabel('Rating')
      plt.title('Release Year vs. Rating')
      plt.show()
```

Release Year vs. Rating
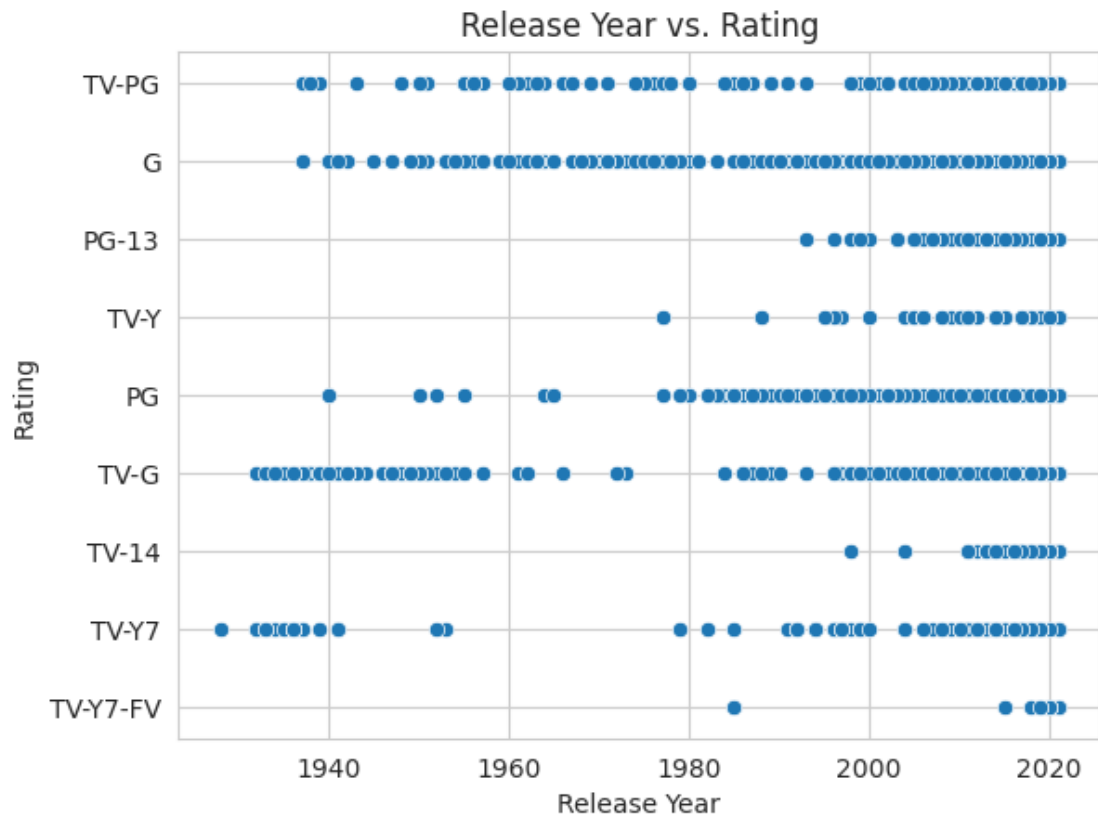
```
disney_data.groupby('release_year')['title'].count().plot(kind='line')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.title('Number of Titles Released per Year')
plt.show()
```

Number of Titles Released per Year