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


from google.colab import files
uploaded = files.upload()
```

```
Choose Files No file chosen  Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
```

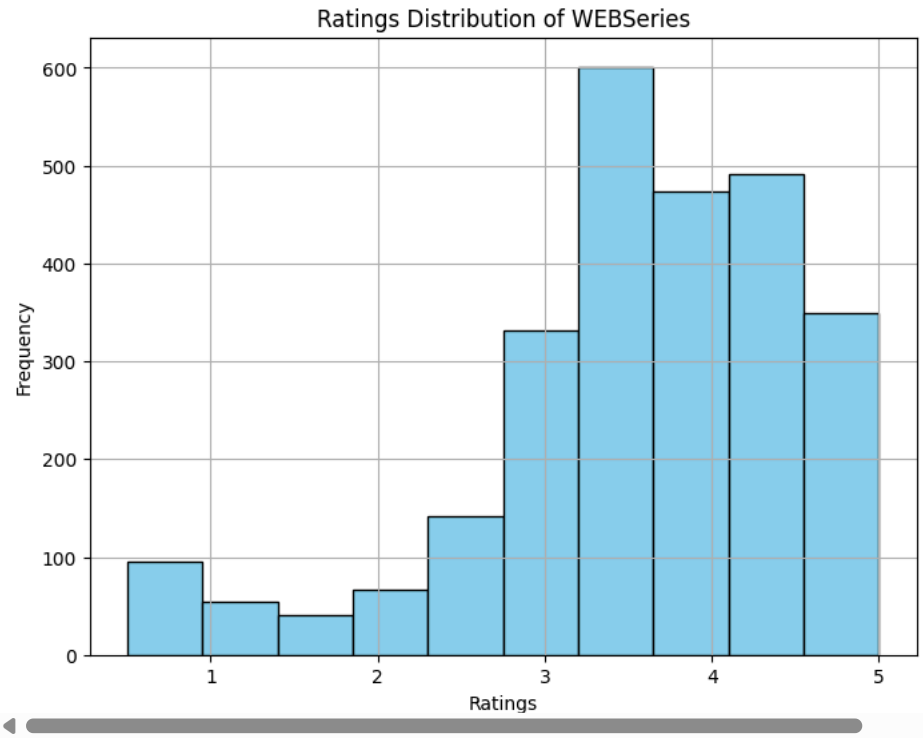
```
data = pd.read_csv("webseries (2).csv")
print(data)
```

```
      Unnamed: 0                   series_name  \
0              1                      Commando
1              2  The Kashmir Files: Unreported
2              3       Made in Heaven Season 2
3              7           The Jengaburu Curse
4              9                    Zombieverse
...          ...                           ...
2639        4992                       Reacher
2640        4993          Rocket Boys Season 1
2641        4994                      Suspicion
2642        4995       The Great Indian Murder
2643        4999         Raising Dion Season 2

                                        director_name             genre  \
0                                 Vipul Amrutlal Shah  2023, 4 Episodes
1                                    Vivek Agnihotri  2023, 7 Episodes
2     Zoya Akhtar, Nitya Mehra, Alankrita Shrivastav...  2023, 7 Episodes
3                                   Nila Madhab Panda  2023, 7 Episodes
4                       No description available  2023, 8 Episodes
...                                                ...              ...
2639  Norberto Barba, Christine Moore, Lin Oeding, M...  2022, 8 Episodes
2640                                     Abhay Pannu  2022, 8 Episodes
2641                    Chris Long, Stefan Schwartz  2022, 8 Episodes
2642                                Tigmanshu Dhulia  2022, 9 Episodes
2643                    No description available  2022, 8 Episodes

      ratings                                      performers  \
0         3.5  Prem Pariijaa, Adah Sharma, Shreya Chaudhary, ...
1         1.3                   Vivek Agnihotri, Pallavi Joshi
2         2.9  Arjun Mathur, Sobhita Dhulipala, Jim Sarbh, Sh...
3         3.9  Faria Abdullah, Nassar, Makrand Deshpande, Sud...
4         0.8  Lee Si-young, Ro Hong-chul, Park Na-rae, Dex, ...
...       ...                                              ...
2639      3.4  Alan Ritchson, Malcolm Goodwin, Willa Fitzgera...
2640      4.1  Ishwak Singh, Jim Sarbh, Regina Cassandra, Raj...
2641      3.5  Uma Thurman, Kunal Nayyar, Noah Emmerich, Geor...
2642      2.9  Richa Chadha, Pratik Gandhi, Ashutosh Rana, Ra...
2643      3.7  Alisha Wainwright, Ja'Siah Young, Jazmyn Simon...

        ott_platform
0     Disney+ Hotstar
1                Zee5
2         Prime Video
3             SonyLIV
4             Netflix
...               ...
2639      Prime Video
2640          SonyLIV
2641        Apple TV+
2642  Disney+ Hotstar
2643          Netflix

[2644 rows x 7 columns]
```

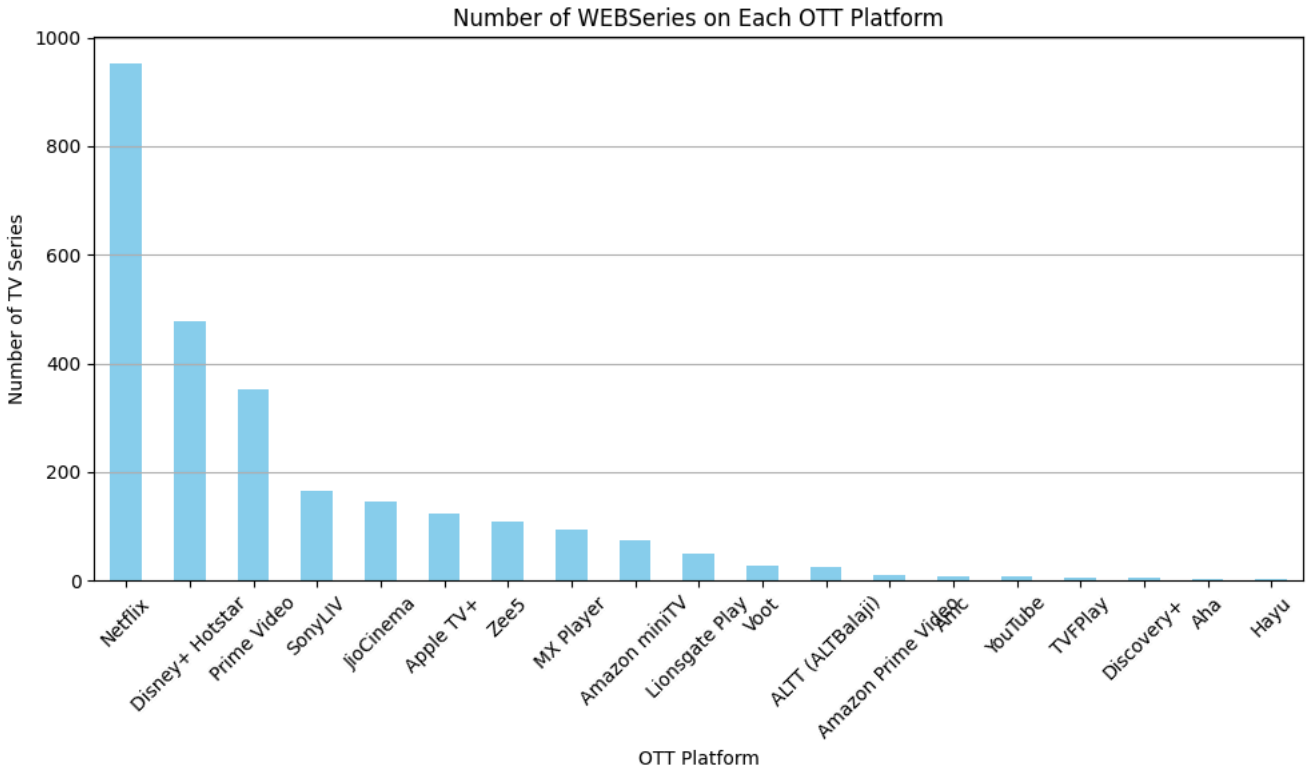
```
print(data.head(15))          #printing 15 lines from start and end
print(data.tail(15))
```

```
2632                          No description available   2022, 7 Episodes
2633                                Coodie, Chike Ozah   2022, 3 Episodes
2634                                     Sacha Jenkins   2022, 4 Episodes
2635                          No description available   2022, 9 Episodes
2636                     Andrew Woodhouse, Dipesh Mistry   2022, 9 Episodes
2637                                   Ritam Srivastav   2022, 9 Episodes
2638                                       Jo Hyun-tak  2022, 10 Episodes
2639  Norberto Barba, Christine Moore, Lin Oeding, M...   2022, 8 Episodes
2640                                       Abhay Pannu   2022, 8 Episodes
2641                          Chris Long, Stefan Schwartz   2022, 8 Episodes
2642                                  Tigmanshu Dhulia   2022, 9 Episodes
2643                          No description available   2022, 8 Episodes

      ratings                                          performers  \
2629      3.0  Conleth Hill, Lesley Manville, Alexandros Logo...
2630      3.3  Huma Qureshi, Avantika Dassani, Parambrata Cha...
2631      4.5  Adam Scott, Britt Lower, John Turturro, Christ...
2632      4.5  Steve Carell, John Malkovich, Ben Schwartz, Di...
2633      5.0                        Kanye West, Yasiin Bey, Jay Z
2634      1.0  Amanda Seales, Jemele Hill, Margaret Cho, Ibra...
2635      4.5  Anna Chlumsky, Julia Garner, Katie Lowes, Lave...
2636      0.5  Nadim Pedrad, Alex Jayne Go, Amari McCoy, Grac...
2637      3.5  Nikitin Dheer, Kranti Prak, Soundarya Sharma, ...
2638      3.1  Jung Hae-in, Jisoo, Yoo In-na, Jang Seung-jo, ...
2639      3.4  Alan Ritchson, Malcolm Goodwin, Willa Fitzgera...
2640      4.1  Ishwak Singh, Jim Sarbh, Regina Cassandra, Raj...
2641      3.5  Uma Thurman, Kunal Nayyar, Noah Emmerich, Geor...
2642      2.9  Richa Chadha, Pratik Gandhi, Ashutosh Rana, Ra...
2643      3.7  Alisha Wainwright, Ja'Siah Young, Jazmyn Simon...

        ott_platform
2629          SonyLIV
2630             Zee5
2631         Apple TV+
2632           Netflix
2633           Netflix
2634         JioCinema
2635           Netflix
2636         Apple TV+
2637         MX Player
2638  Disney+ Hotstar
2639      Prime Video
2640          SonyLIV
2641         Apple TV+
2642  Disney+ Hotstar
2643           Netflix
```

```
# Plotting the ratings distribution using a histogram
plt.figure(figsize=(8, 6))
plt.hist(data['ratings'], bins=10, color='skyblue', edgecolor='black')
plt.title('Ratings Distribution of WEBSeries')
plt.xlabel('Ratings')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



```
#a visualization to compare the number of WEBseries available on each OTT platform using a bar chart
# we will count the occurrences of each platform and then plot the results.

# Count the number of TV series on each OTT platform
platform_counts = data['ott_platform'].value_counts()

# Plotting the bar chart
plt.figure(figsize=(10, 6))
platform_counts.plot(kind='bar', color='skyblue')
plt.title('Number of WEBSeries on Each OTT Platform')
plt.xlabel('OTT Platform')
plt.ylabel('Number of TV Series')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.grid(axis='y')  # Show gridlines on the y-axis
plt.tight_layout()  # Adjust layout to prevent clipping of labels
plt.show()
```




we can see that most of the webseries is available on netflix thus it will be beneficial for the customers to subscribe netflix

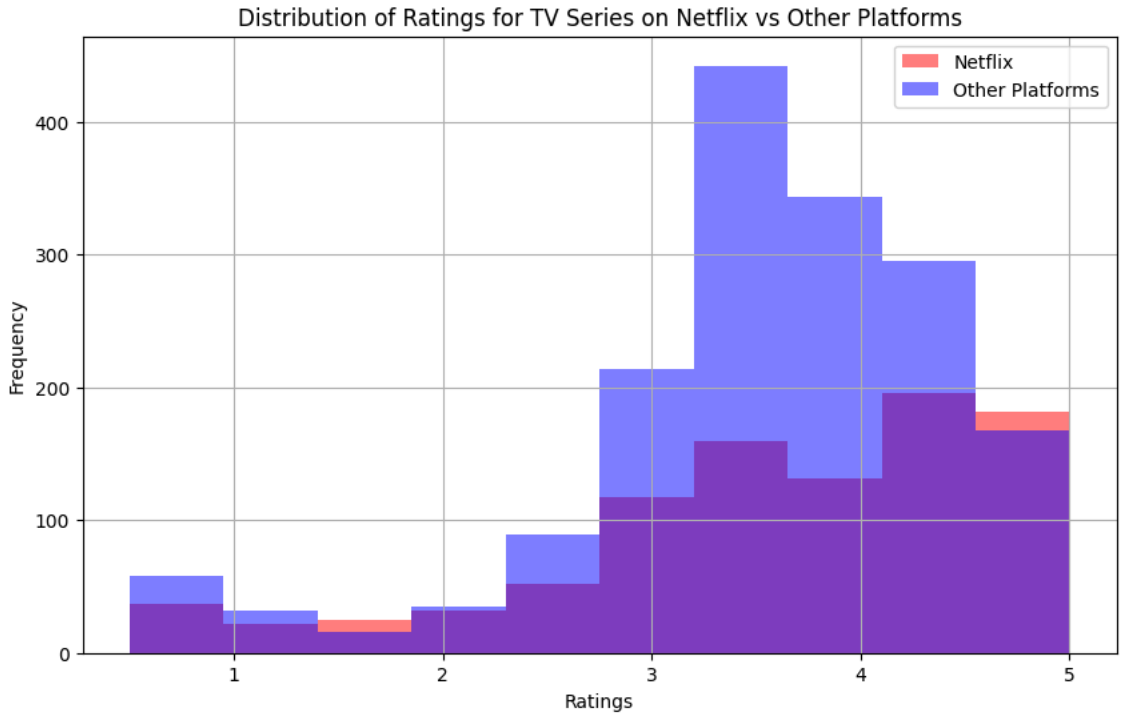
Since Netflix seems to have the highest number of TV series in your dataset, it would be interesting to visualize the distribution of ratings for TV series available on Netflix compared to other platforms

**RATING DISTRIBUTION of NETFLIX and other OTT platforms**

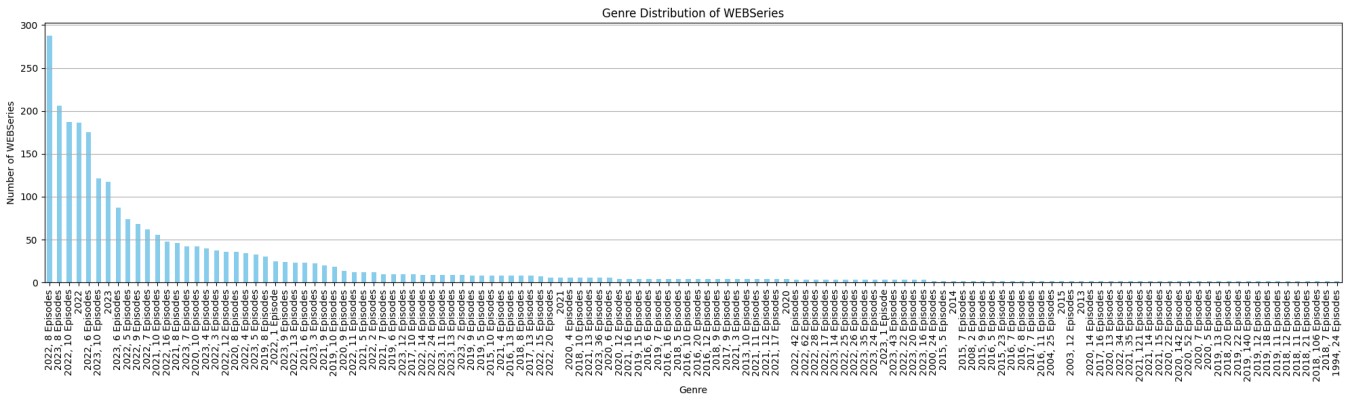
```
# Filter the DataFrame to include only TV series available on Netflix
netflix_series = data[data['ott_platform'] == 'Netflix']

# Filter the DataFrame to include TV series not available on Netflix
other_series = data[data['ott_platform'] != 'Netflix']

# Plotting the distribution of ratings for Netflix series and other series
plt.figure(figsize=(10, 6))

# Plotting Netflix series ratings
plt.hist(netflix_series['ratings'], bins=10, color='red', alpha=0.5, label='Netflix')

# Plotting ratings of series on other platforms
plt.hist(other_series['ratings'], bins=10, color='blue', alpha=0.5, label='Other Platforms')

plt.title('Distribution of Ratings for TV Series on Netflix vs Other Platforms')
plt.xlabel('Ratings')
plt.ylabel('Frequency')
```

```
plt.legend()
plt.grid(True)
plt.show()
```

```
#netflixseries vs rating
```



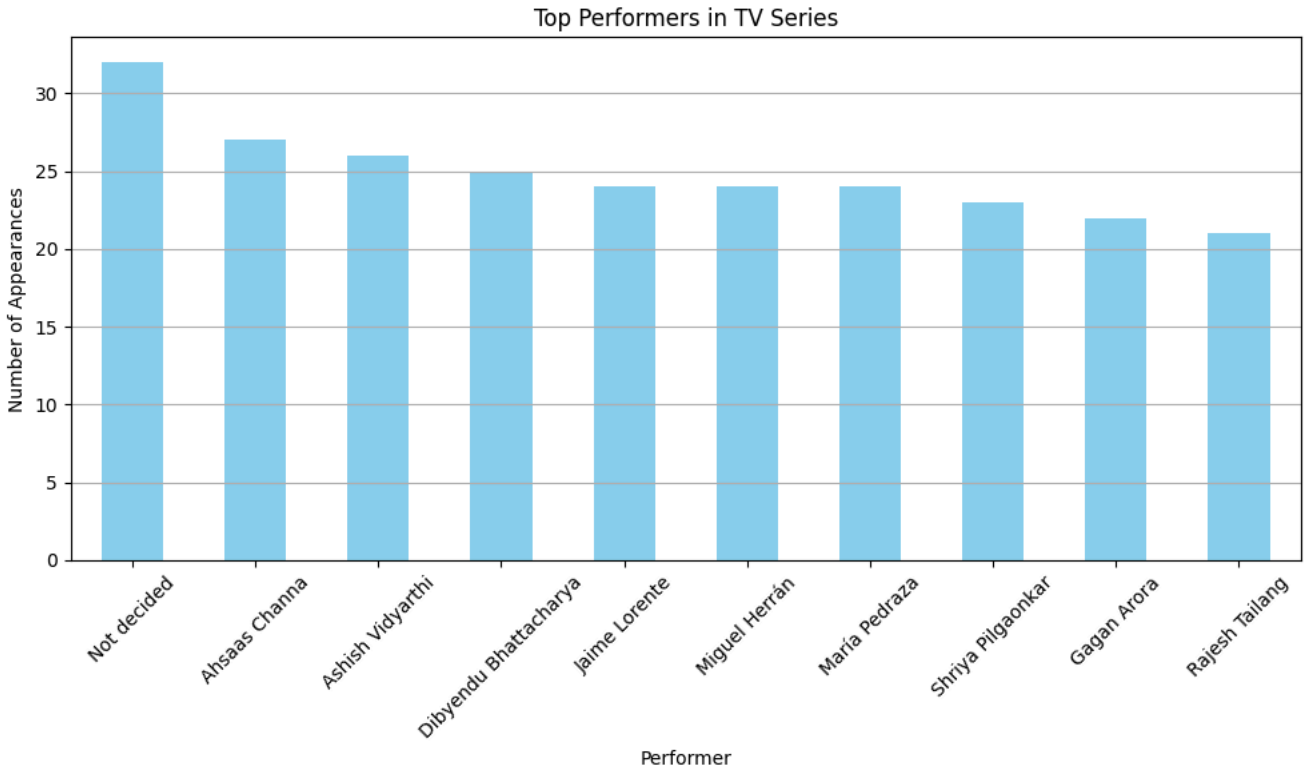

```
# Count the number of WEB series in each genre
genre_counts = data['genre'].value_counts()

# Plotting the genre distribution using a bar chart
plt.figure(figsize=(20, 6))
genre_counts.plot(kind='bar', color='skyblue')
plt.title('Genre Distribution of WEBSeries')
plt.xlabel('Genre')
plt.ylabel('Number of WEBSeries')
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.grid(axis='y')  # Show gridlines on the y-axis
plt.tight_layout()  # Adjust layout to prevent clipping of labels
plt.show()
```




Lesser the number of episodes , more impactful the series is

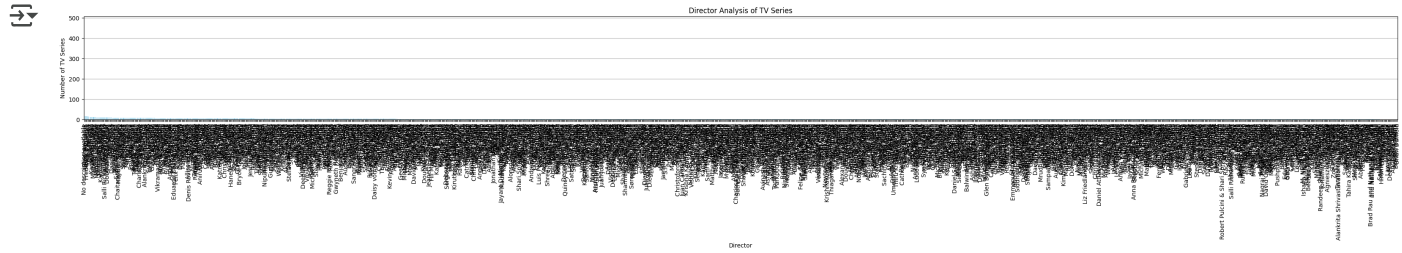
**TOP PERFORMERS**

```
performers = data['performers'].str.split(', ', expand=True).stack()

# Count the occurrences of each performer
top_performers = performers.value_counts().head(10)  # Adjust the number based on your preference

# Plotting the top performers
plt.figure(figsize=(10, 6))
top_performers.plot(kind='bar', color='skyblue')
plt.title('Top Performers in TV Series')
plt.xlabel('Performer')
plt.ylabel('Number of Appearances')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.grid(axis='y')  # Show gridlines on the y-axis
plt.tight_layout()  # Adjust layout to prevent clipping of labels
plt.show()
```

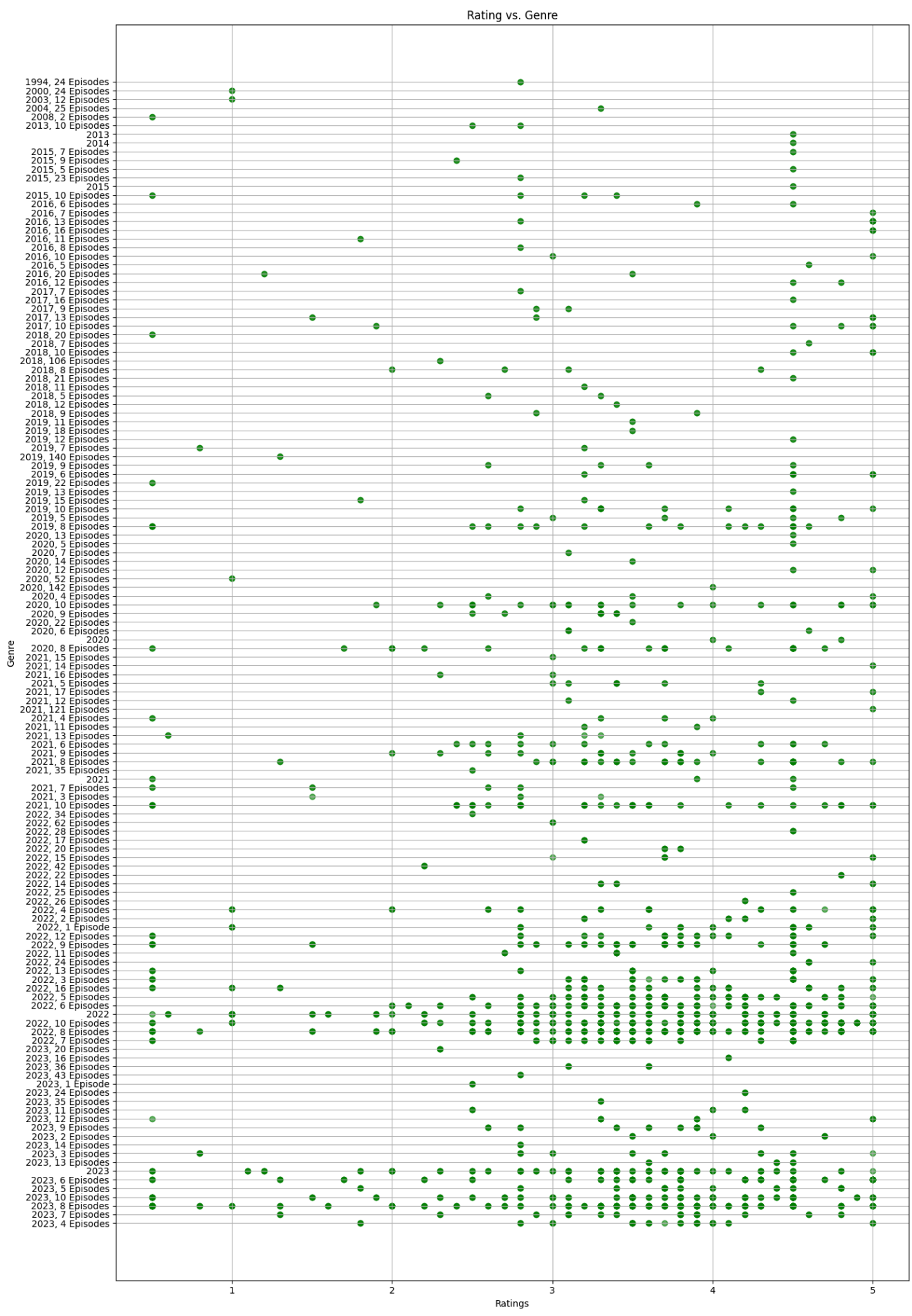



**DIRECTOR ANALYSIS**

```
directors = data['director_name'].str.split(', ', expand=True).stack()

# Count the occurrences of each director
director_counts = directors.value_counts()

# Plotting the director analysis using a bar chart
plt.figure(figsize=(32, 6))
director_counts.plot(kind='bar', color='skyblue')
plt.title('Director Analysis of TV Series')
plt.xlabel('Director')
plt.ylabel('Number of TV Series')
plt.xticks(rotation=90)
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```




```
# Plotting rating vs. genre using a scatter plot
plt.figure(figsize=(14, 20))
plt.scatter(data['ratings'], data['genre'], color='green', alpha=0.6)
plt.title('Rating vs. Genre')
plt.xlabel('Ratings')
plt.ylabel('Genre')
plt.grid(True)
plt.tight_layout()
plt.show()
```


Rating vs. Genre
Genre
Ratings
1994, 24 Episodes
2000, 24 Episodes
2003, 12 Episodes
2004, 25 Episodes
2008, 2 Episodes
2013, 10 Episodes
2013
2014
2015, 7 Episodes
2015, 9 Episodes
2015, 5 Episodes
2015, 23 Episodes
2015
2015, 10 Episodes
2016, 6 Episodes
2016, 7 Episodes
2016, 13 Episodes
2016, 16 Episodes
2016, 11 Episodes
2016, 8 Episodes
2016, 10 Episodes
2016, 5 Episodes
2016, 20 Episodes
2016, 12 Episodes
2017, 7 Episodes
2017, 16 Episodes
2017, 9 Episodes
2017, 13 Episodes
2017, 10 Episodes
2018, 20 Episodes
2018, 7 Episodes
2018, 10 Episodes
2018, 106 Episodes
2018, 8 Episodes
2018, 21 Episodes
2018, 11 Episodes
2018, 5 Episodes
2018, 12 Episodes
2018, 9 Episodes
2019, 11 Episodes
2019, 18 Episodes
2019, 12 Episodes
2019, 7 Episodes
2019, 140 Episodes
2019, 9 Episodes
2019, 6 Episodes
2019, 22 Episodes
2019, 13 Episodes
2019, 15 Episodes
2019, 10 Episodes
2019, 5 Episodes
2019, 8 Episodes
2020, 13 Episodes
2020, 5 Episodes
2020, 7 Episodes
2020, 14 Episodes
2020, 12 Episodes
2020, 52 Episodes
2020, 142 Episodes
2020, 4 Episodes
2020, 10 Episodes
2020, 9 Episodes
2020, 22 Episodes
2020, 6 Episodes
2020
2020, 8 Episodes
2021, 15 Episodes
2021, 14 Episodes
2021, 16 Episodes
2021, 5 Episodes
2021, 17 Episodes
2021, 12 Episodes
2021, 121 Episodes
2021, 4 Episodes
2021, 11 Episodes
2021, 13 Episodes
2021, 6 Episodes
2021, 9 Episodes
2021, 8 Episodes
2021, 35 Episodes
2021
2021, 7 Episodes
2021, 3 Episodes
2021, 10 Episodes
2022, 34 Episodes
2022, 62 Episodes
2022, 28 Episodes
2022, 17 Episodes
2022, 20 Episodes
2022, 15 Episodes
2022, 42 Episodes
2022, 22 Episodes
2022, 14 Episodes
2022, 25 Episodes
2022, 26 Episodes
2022, 4 Episodes
2022, 2 Episodes
2022, 1 Episode
2022, 12 Episodes
2022, 9 Episodes
2022, 11 Episodes
2022, 24 Episodes
2022, 13 Episodes
2022, 3 Episodes
2022, 16 Episodes
2022, 5 Episodes
2022, 6 Episodes
2022
2022, 10 Episodes
2022, 8 Episodes
2022, 7 Episodes
2023, 20 Episodes
2023, 16 Episodes
2023, 36 Episodes
2023, 43 Episodes
2023, 1 Episode
2023, 24 Episodes
2023, 35 Episodes
2023, 11 Episodes
2023, 12 Episodes
2023, 9 Episodes
2023, 2 Episodes
2023, 14 Episodes
2023, 3 Episodes
2023, 13 Episodes
2023
2023, 6 Episodes
2023, 5 Episodes
2023, 10 Episodes
2023, 8 Episodes
2023, 7 Episodes
2023, 4 Episodes
1
2
3
4
5

**Rating trends over time**

```
#as there is no release year present in the dataset we cannot tell that which is the best year for the OTT platforms
```

**PLATFORM VS GENRE**

```
platform_genre_counts = data.groupby(['ott_platform', 'genre']).size().unstack(fill_value=0)

# Plotting platform vs. genre using a stacked bar chart
plt.figure(figsize=(12, 8))
platform_genre_counts.plot(kind='bar', stacked=True, cmap='tab20')
plt.title('Platform vs. Genre')
plt.xlabel('OTT Platform')
plt.ylabel('Number of TV Series')
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.legend(title='Genre', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```