Netflix Data Analysis

Ishita Mishra

Prakhar Tripathi

Netflix Data Analysis

Description:

We're taking a close look at a dataset of Netflix Movies and TV shows to learn interesting things and solve problems. It's like a detective work for data! We're checking how the data is spread, finding patterns, and figuring out if anything is missing. This helps us understand the dataset better and make smart decisions based on what we find.

Data Preprocessing & Exploration

- We're planning to analyze a dataset using Pandas, Matplotlib, and Seaborn. These tools will assist us in cleaning up the data and creating visualizations. Think of it like using special tools to tidy up and then draw pictures of the information to understand it better.

- Importing the required packages :

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

Importing the Dataset:

```
df = pd.read_csv("C:/Users/ishita mishra/Desktop/Netflix/Netflix Dataset.csv")
```

- Displaying Top and Bottom 10 Values of the Dataset:

Top 10 Values -- print(df.head(10))

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
1	s2	Movie	07:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	R	78 min	Horror Movies, International Movies	When an army recruit is found dead, his fellow...
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...	In a postapocalyptic world, rag-doll robots hi...
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	PG-13	123 min	Dramas	A brilliant group of students become card-coun...
5	s6	TV Show	46	Serdar Akar	Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan...	Turkey	July 1, 2017	TV-MA	1 Season	International TV Shows, TV Dramas, TV Mysteries	A genetics professor experiments with a treatm...
6	s7	Movie	122	Yasir Al Yasiri	Amina Khalil, Ahmed Dawood, Tarek Lotfy, Ahmed...	Egypt	June 1, 2020	TV-MA	95 min	Horror Movies, International Movies	After an awful accident, a couple admitted to ...
7	s8	Movie	187	Kevin Reynolds	Samuel L. Jackson, John Heard, Kelly Rowan, Cl...	United States	November 1, 2019	R	119 min	Dramas	After one of his high school students attacks ...
8	s9	Movie	706	Shravan Kumar	Divya Dutta, Atul Kulkarni, Mohan Agashe, Anup...	India	April 1, 2019	TV-14	118 min	Horror Movies, International Movies	When a doctor goes missing, his psychiatrist w...
9	s10	Movie	1920	Vikram Bhatt	Rajneesh Duggal, Adah Sharma, Indraneil Sengup...	India	December 15, 2017	TV-MA	143 min	Horror Movies, International Movies, Thrillers	An architect and his wife move into a castle t...

- Bottom 10 Values -- print(df.tail(10))

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
7781	s7780	TV Show	Zona Rosa	NaN	Manu NNa, Ana Julia Yeyé, Ray Contreras, Pablo...	Mexico	November 26, 2019	TV-MA	1 Season	International TV Shows, Spanish-Language TV Sh...	An assortment of talent takes the stage for a ...
7782	s7781	Movie	Zoo	Shlok Sharma	Shashank Arora, Shweta Tripathi, Rahul Kumar, ...	India	July 1, 2018	TV-MA	94 min	Dramas, Independent Movies, International Movies	A drug dealer starts having doubts about his t...
7783	s7782	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
7784	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	TV-MA	99 min	Dramas, International Movies	When Lebanon's Civil War deprives Zozo of his ...
7785	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...
7786	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	TV-MA	44 min	Documentaries, International Movies, Music & M...	In this documentary, South African rapper Nast...
7787	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	TV-PG	1 Season	International TV Shows, Reality TV	Dessert wizard Adriano Zumbo looks for the nex...
7788	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	TV-MA	90 min	Documentaries, Music & Musicals	This documentary delves into the mystique behi...

- Understanding the structure of the data -- print(df.info())

In the exploratory phase, it has been observed that the dataset does not contain any null values. As a result, there is no immediate need for data imputation or handling missing values

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7789 entries, 0 to 7788
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   Show_Id       7789 non-null   object
 1   Category      7789 non-null   object
 2   Title         7789 non-null   object
 3   Director      5401 non-null   object
 4   Cast          7071 non-null   object
 5   Country       7282 non-null   object
 6   Release_Date  7779 non-null   object
 7   Rating        7782 non-null   object
 8   Duration      7789 non-null   object
 9   Type          7789 non-null   object
 10  Description   7789 non-null   object
dtypes: object(11)
memory usage: 669.5+ KB
```

- For Stastical Analysis -- print(df.describe())

	Show_Id	Category	Title	Director	Cast	Country	Release_Date	Rating	Duration	Type	Description
count	7789	7789	7789	5401	7071	7282	7779	7782	7789	7789	7789
unique	7787	2	7787	4050	6831	681	1565	14	216	492	7769
top	s6621	Movie	The Lost Okoroshi	Raúl Campos, Jan Suter	David Attenborough	United States	January 1, 2020	TV-MA	1 Season	Documentaries	Multiple women report their husbands as missin...
freq	2	5379	2	18	18	2556	118	2865	1608	334	3

- Checking the column names - print(df.columns)

```
Index(['Show_Id', 'Category', 'Title', 'Director', 'Cast', 'Country',
       'Release_Date', 'Rating', 'Duration', 'Type', 'Description'],
      dtype='object')
```

- Fetching the year from Releasing Date –

```
df["Releasing_Year"] = pd.DatetimeIndex(df["Release_Date"]).year
print(df["Releasing_Year"])
```

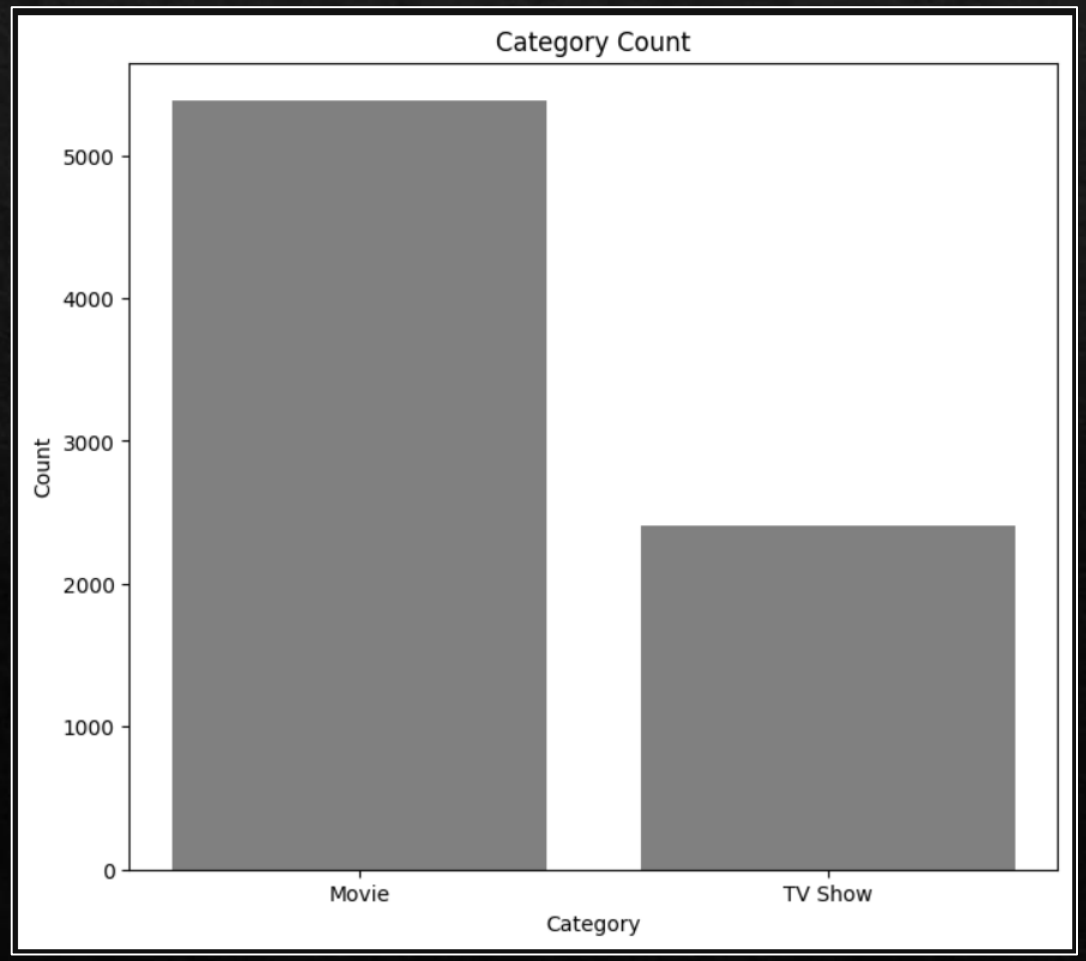
```
0       2020.0
1       2016.0
2       2018.0
3       2017.0
4       2020.0
         ...
7784    2020.0
7785    2019.0
7786    2020.0
7787    2020.0
7788    2020.0
Name: Releasing_Year, Length: 7789, dtype: float64
```

Exploratory Data Analysis & Visualization

- Conducting an analysis to determine the prevalence of content categories on Netflix, with the goal of identifying the most prominent genres within the platform's offerings.

```
print(df["Category"].value_counts())
plt.figure(figsize=(8,7))
Category = ["Movie","TV Show"]
Value = df["Category"].value_counts()
plt.bar(Category,Value, color = "grey")
plt.xlabel("Category")
plt.ylabel("Count")
plt.title("Category Count")
plt.show()
```

```
Category
Movie      5379
TV Show    2410
Name: count, dtype: int64
```

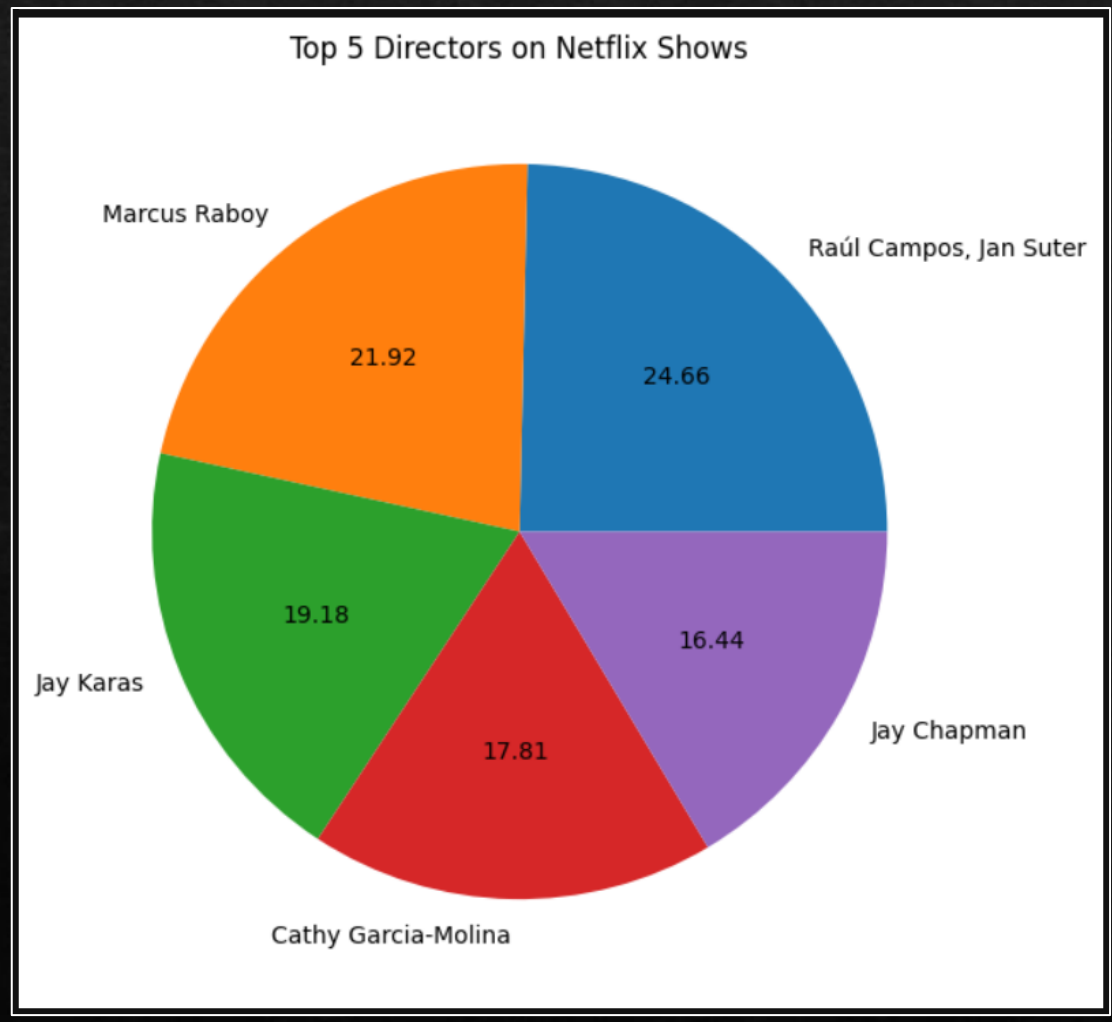


- Through visualization, it is evident that the 'Movies' category on Netflix is twice as prevalent compared to 'TV Shows,'

- Conducting an analysis to identify the top five directors who have produced the maximum number of movies.

```
print(df["Director"].value_counts()[:5])
plt.figure(figsize=(10,7))
plt.title("Top 5 Directors on Netflix Shows")
Directors = df["Director"].value_counts()[:5]
plt.pie(Directors,autopct="%.2f", labels=Directors.index)
plt.show()
```

```
Director
Raúl Campos, Jan Suter    18
Marcus Raboy              16
Jay Karas                 14
Cathy Garcia-Molina       13
Jay Chapman               12
Name: count, dtype: int64
```

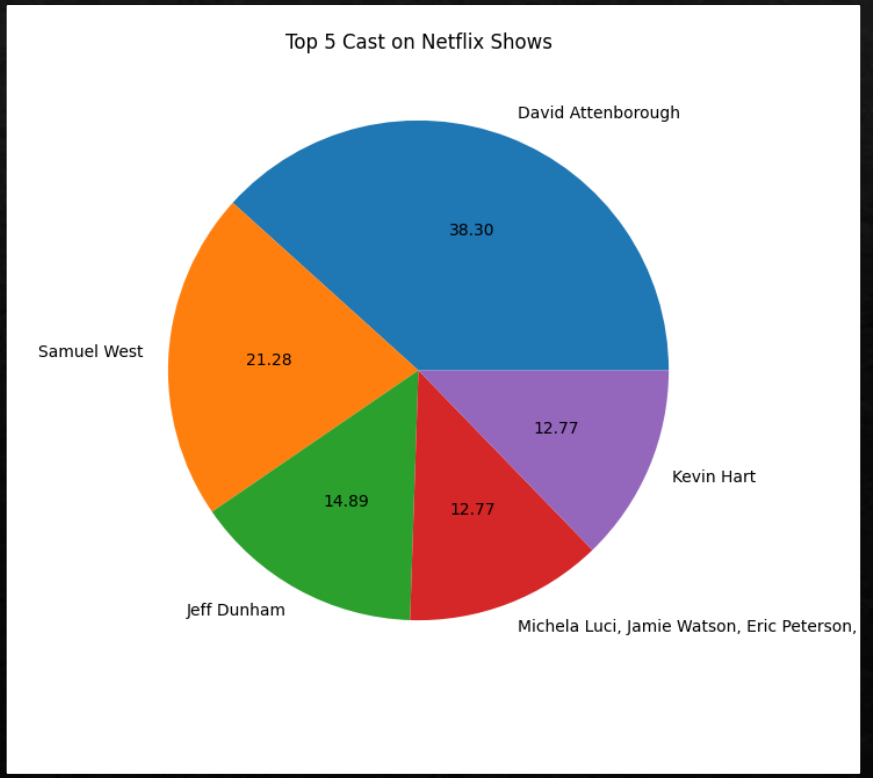


- Exploring the dataset to identify the top five cast members who have been featured in the highest number of movies and TV shows on Netflix.

```
print(df["Cast"].value_counts()[:5])
```

```
Cast
David Attenborough                                                                  18
Samuel West                                                                         10
Jeff Dunham                                                                          7
Michela Luci, Jamie Watson, Eric Peterson, Anna Claire Bartlam, Nicolas Aqui, Cory Doran, Julie Lemieux, Derek McGrath     6
Kevin Hart                                                                           6
Name: count, dtype: int64
```

```
plt.figure(figsize=(10,7))
plt.title("Top 5 Cast on Netflix Shows")
Cast = df["Cast"].value_counts()[:5]
plt.pie(Cast,autopct="%.2f", labels=Cast.index)
plt.show()
```

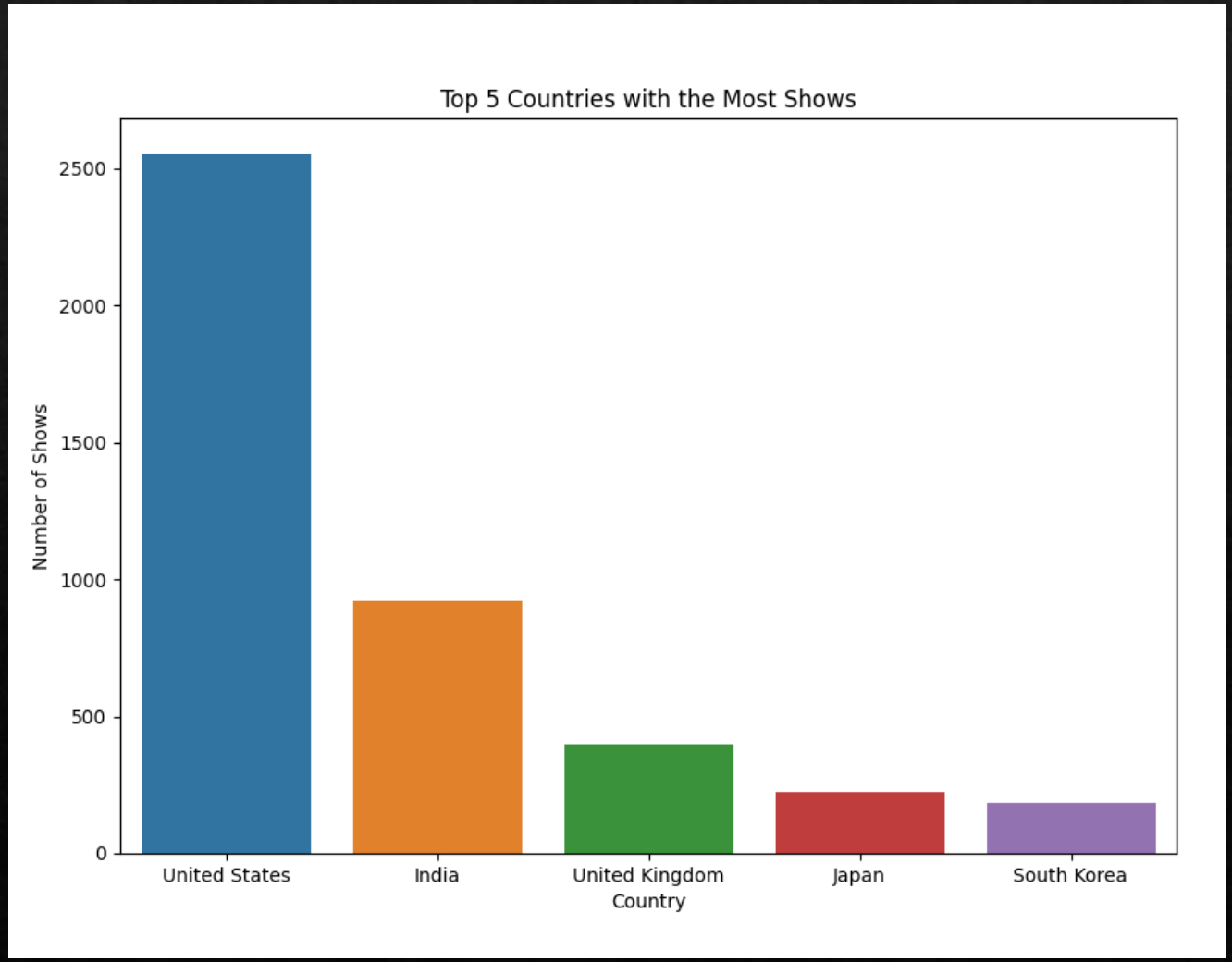


- Analyzing the dataset to pinpoint the top five countries contributing the highest number of shows on Netflix, providing insights into the most prolific content-producing nations on the platform.

```
print(df["Country"].value_counts()[0:5])
```

```
plt.figure(figsize=(10,7))
top_countries = df["Country"].value_counts()[:5]
sns.barplot(x=top_countries.index, y=top_countries.values)
plt.title('Top 5 Countries with the Most Shows')
plt.xlabel('Country')
plt.ylabel('Number of Shows')
plt.show()
```

```
Country
United States     2556
India              923
United Kingdom     397
Japan              226
South Korea        183
Name: count, dtype: int64
```

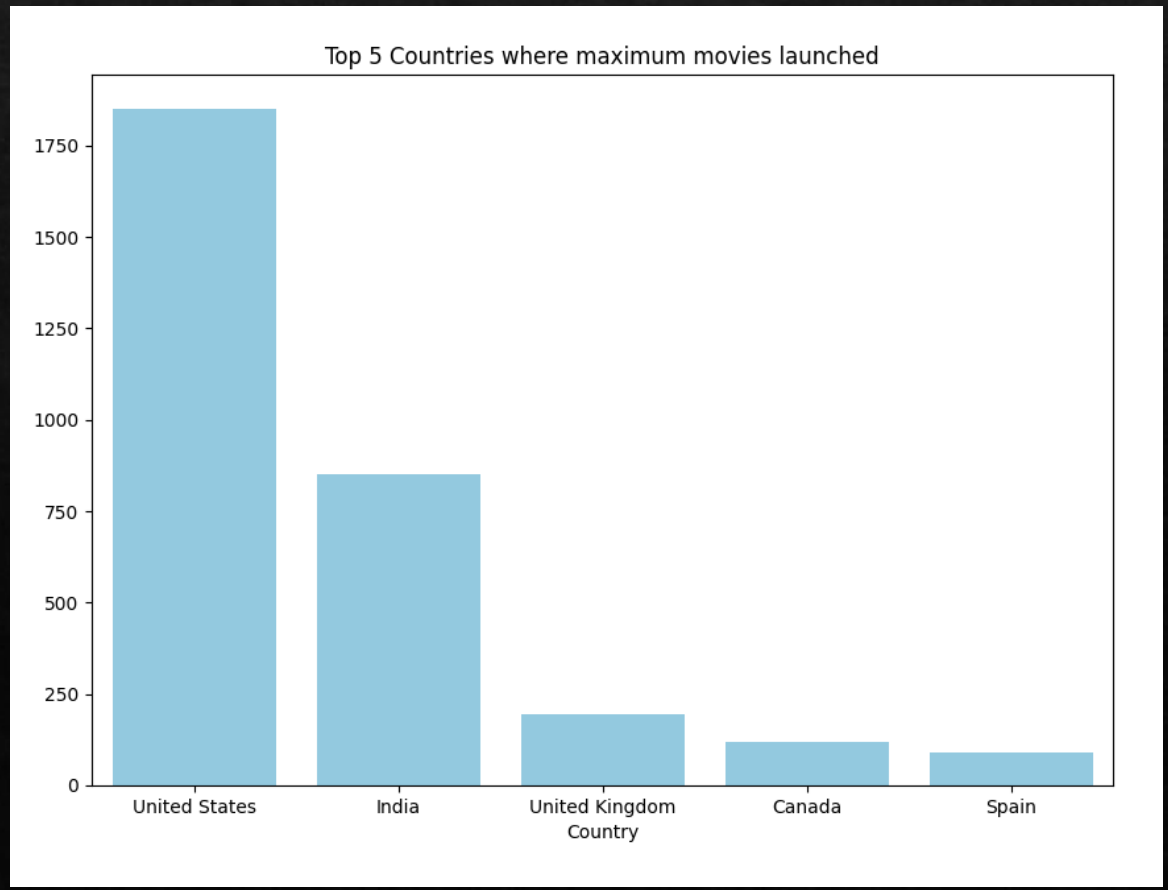


- Conducting an analysis to determine the top five countries with the highest production of movies on Netflix, offering insights into the leading contributors to the platform's movie content.
- To study movies on Netflix, we're creating a new dataset that only has information about movies. This will make it easier for us to analyze and find interesting things about movies specifically.

```
Movie_df = df[df["Category"] == "Movie"]
print(Movie_df["Country"].value_counts()[0:5])
```

```
Country
United States     1851
India              852
United Kingdom     193
Canada             118
Spain               89
Name: count, dtype: int64
```

```
plt.figure(figsize=(10,7))
plt.title("Top 5 Countries where maximum movies launched")
top_countries = Movie_df["Country"].value_counts()[:5]
sns.barplot(x=top_countries.index, y=top_countries.values, color="skyblue")
plt.show()
```



- Exploring the countries where the top director produce movies, aiming to identify the geographic locations where influential directors are most active in filmmaking.
- In order to analyze the works of top directors like Raúl Campos and Jan Suter, we are creating a new dataset that exclusively includes movies and TV shows directed by them. This specialized dataset will enable a focused examination of their contributions to the Netflix platform.

```
Director_df = Movie_df[Movie_df["Director"] == "Raúl Campos, Jan Suter"]
print(Director_df["Country"].value_counts())
```

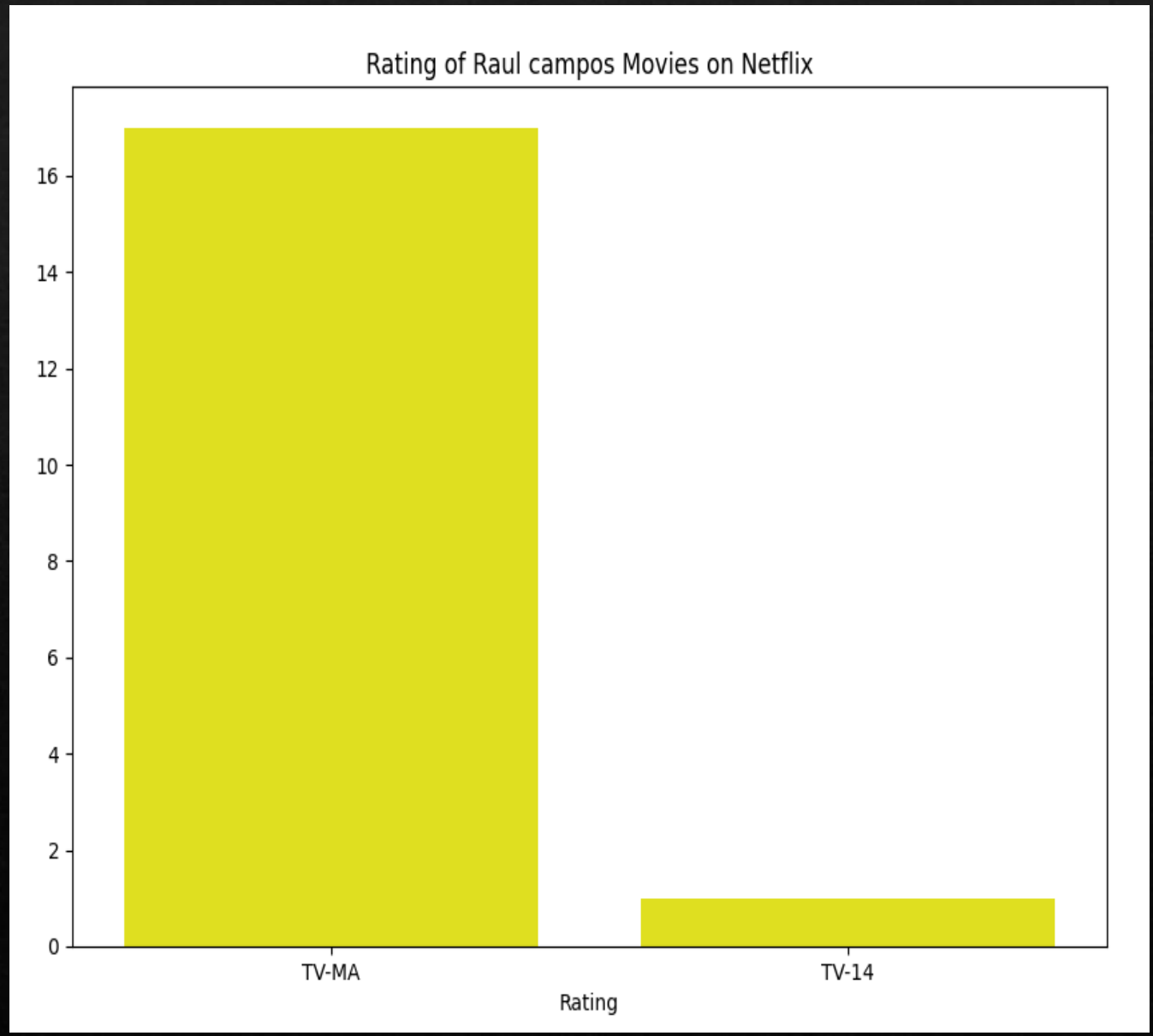
```
Country
Mexico        9
Argentina     5
Colombia      2
Chile         2
Name: count, dtype: int64
```

- Examining the ratings for content directed by top filmmakers Raúl Campos and Jan Suter to assess the audience reception and critical acclaim of their works on Netflix.

```
print(Director_df["Rating"].value_counts()[:5])
```

```
Rating
TV-MA    17
TV-14     1
Name: count, dtype: int64
```

```
plt.figure(figsize=(10,7))
plt.title("Rating of Raul campos Movies on Netflix")
Rating = Director_df["Rating"].value_counts()[:5]
sns.barplot(x = Rating.index, y= Rating.values , color="yellow")
plt.show()
```



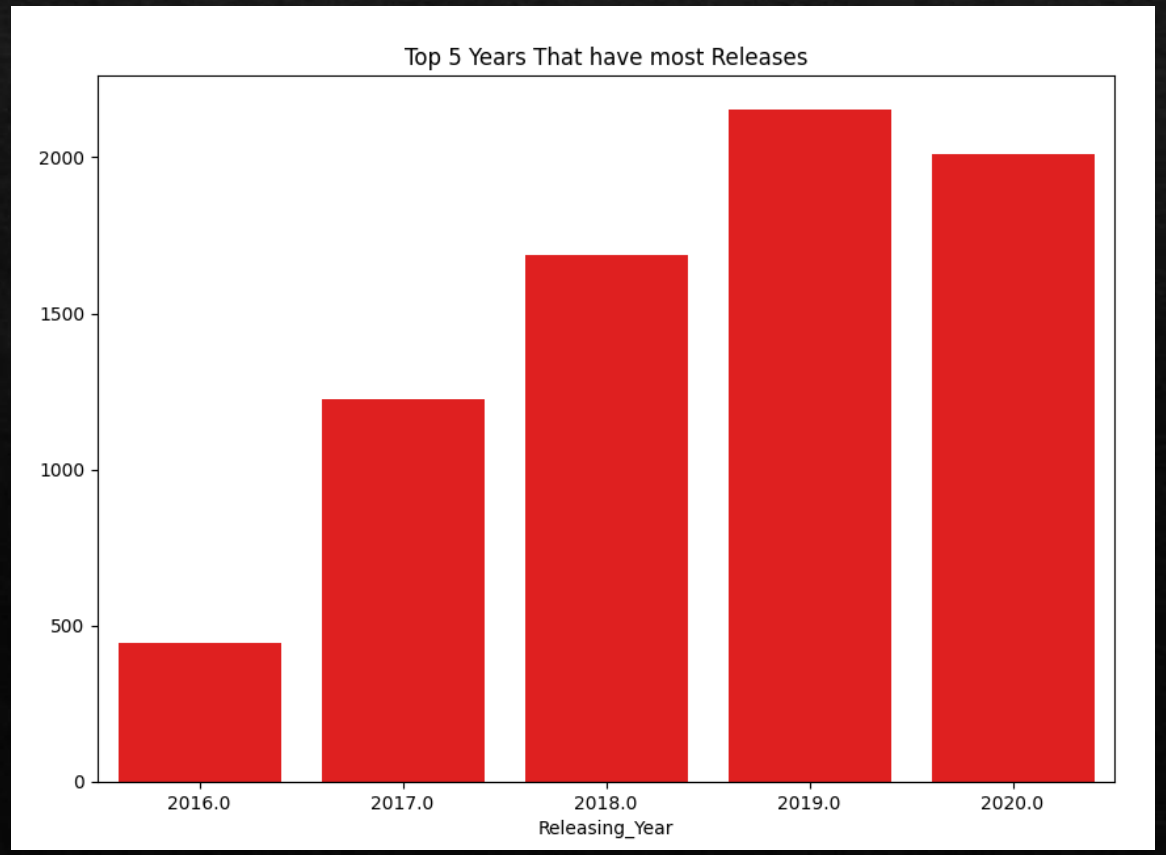
- Assigning content ratings on Netflix, where TV-MA indicates 'Below 17 Age Content' and TV-14 designates 'Mature Only Content,' offering clarity on the age-appropriate categorization of shows on the platform.

- Exploring the dataset to determine the top five most frequent releasing years for content on Netflix, providing insights into the distribution of production years across the platform.

```
print(df["Releasing_Year"].value_counts().head())
```

```
Releasing_Year
2019.0    2154
2020.0    2010
2018.0    1685
2017.0    1225
2016.0     443
Name: count, dtype: int64
```

```
plt.figure(figsize=(10,7))
plt.title("Top 5 Years That have most Releases")
Year= df["Releasing_Year"].value_counts().head()
sns.barplot(x = Year.index, y= Year.values , color="Red")
plt.show()
```

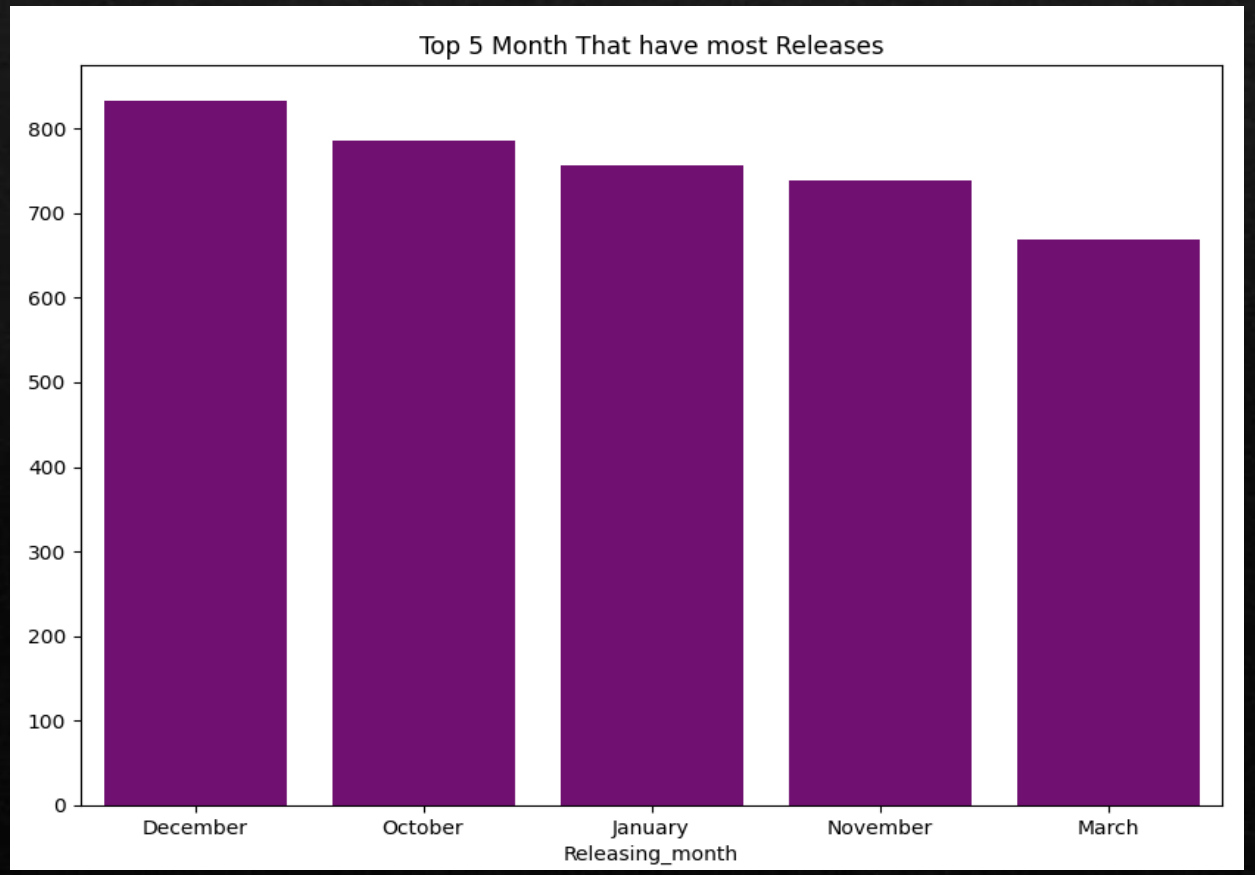


- Based on our analysis, we have determined that the year 2019 witnessed the highest number of content releases on Netflix, indicating a significant concentration of new releases during that specific year.

- Exploring the dataset to determine the top five most frequent releasing Months for content on Netflix, providing insights into the distribution of production months across the platform.
- Extracting the month information from the release date column to gain insights into the distribution of content releases throughout the months, facilitating a temporal analysis of Netflix content.

```
df["Releasing_month"] = pd.DatetimeIndex(df["Release_Date"]).month
df["Releasing_month"] = pd.to_datetime(df["Releasing_month"], format="%m").dt.month_name()
print(df["Releasing_month"].value_counts())
plt.figure(figsize=(10,7))
plt.title("Top 5 Month That have most Releases")
Month= df["Releasing_month"].value_counts().head()
sns.barplot(x = Month.index, y= Month.values , color="purple")
plt.show()
```

```
Releasing_month
December     833
October      785
January      757
November     738
March        669
September    620
August       618
April        602
July         600
May          543
June         542
February     472
Name: count, dtype: int64
```



- Conducting an analysis to identify the top countries contributing the highest number of Netflix shows in the year 2019, revealing insights into the geographic distribution of content during that period..
- In preparation for this analysis, we are creating a new dataset that specifically includes data from the year 2019.

```
Year_df = df[df["Releasing_Year"] == 2019]
print(Year_df["Country"].value_counts()[:5])
```

```
Country
United States     732
India             225
United Kingdom    107
Japan              62
South Korea        58
Name: count, dtype: int64
```

- Exploring the dataset for the year 2019 to identify the top directors who contributed the highest number of Netflix shows during that period.

```
print(Year_df["Director"].value_counts()[:5])
```

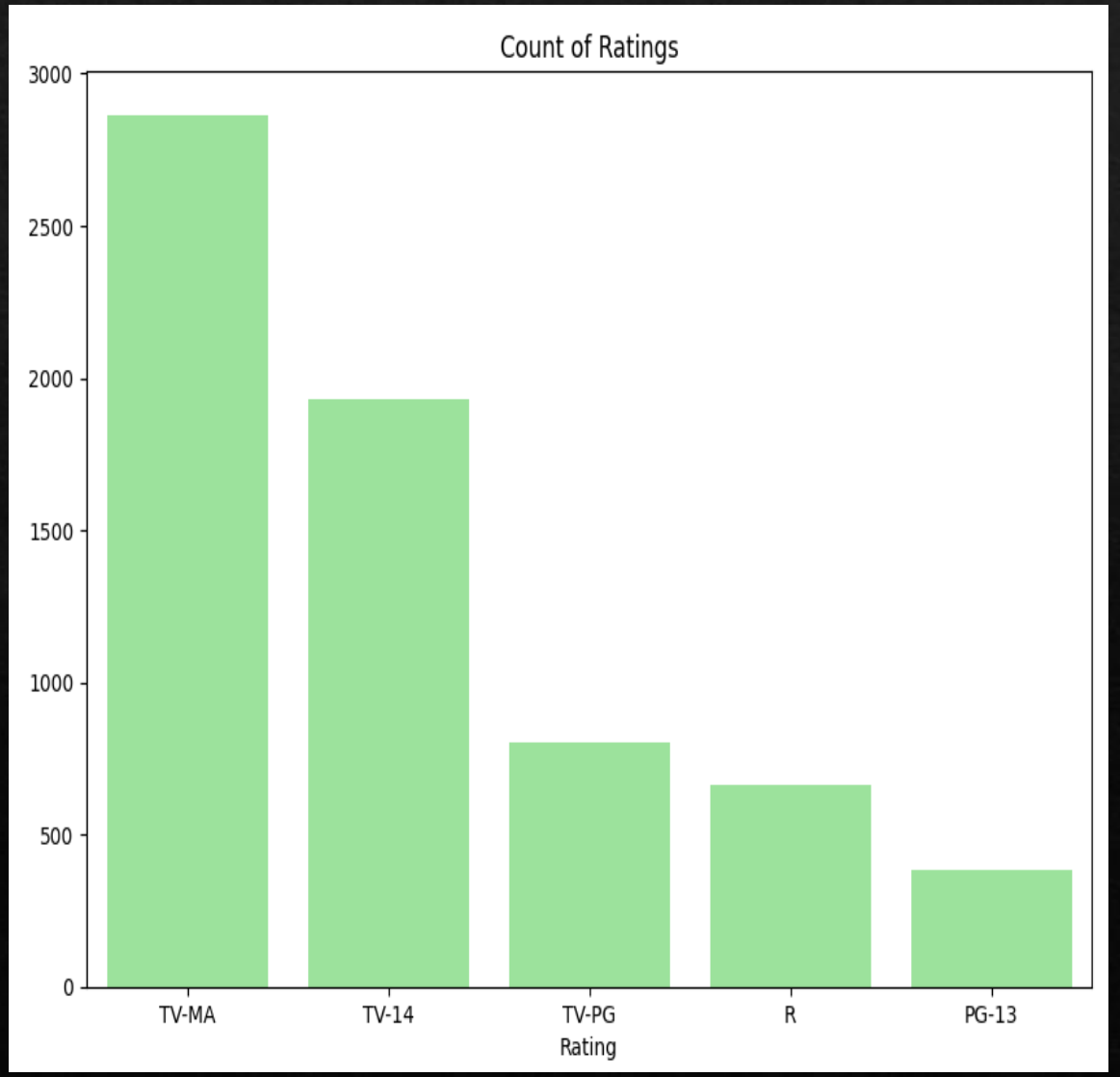
```
Director
Martin Scorsese          7
Cathy Garcia-Molina      7
Steven Spielberg         6
Kunle Afolayan           6
Wenn V. Deramas          6
Name: count, dtype: int64
```

- Calculating the count of ratings for movies and shows on Netflix, providing a quantitative measure of the audience engagement and feedback for the content available on the platform.

```
print(df["Rating"].value_counts())
```

```
Rating
TV-MA       2865
TV-14       1931
TV-PG        806
R            665
PG-13        386
TV-Y         280
TV-Y7        271
PG           247
TV-G         194
NR            84
G             39
TV-Y7-FV       6
UR             5
NC-17          3
Name: count, dtype: int64
```

```
plt.figure(figsize=(10,7))
plt.title("Count of Ratings")
rating= df["Rating"].value_counts().head()
sns.barplot(x = rating.index, y= rating.values , color="lightgreen")
plt.show()
```

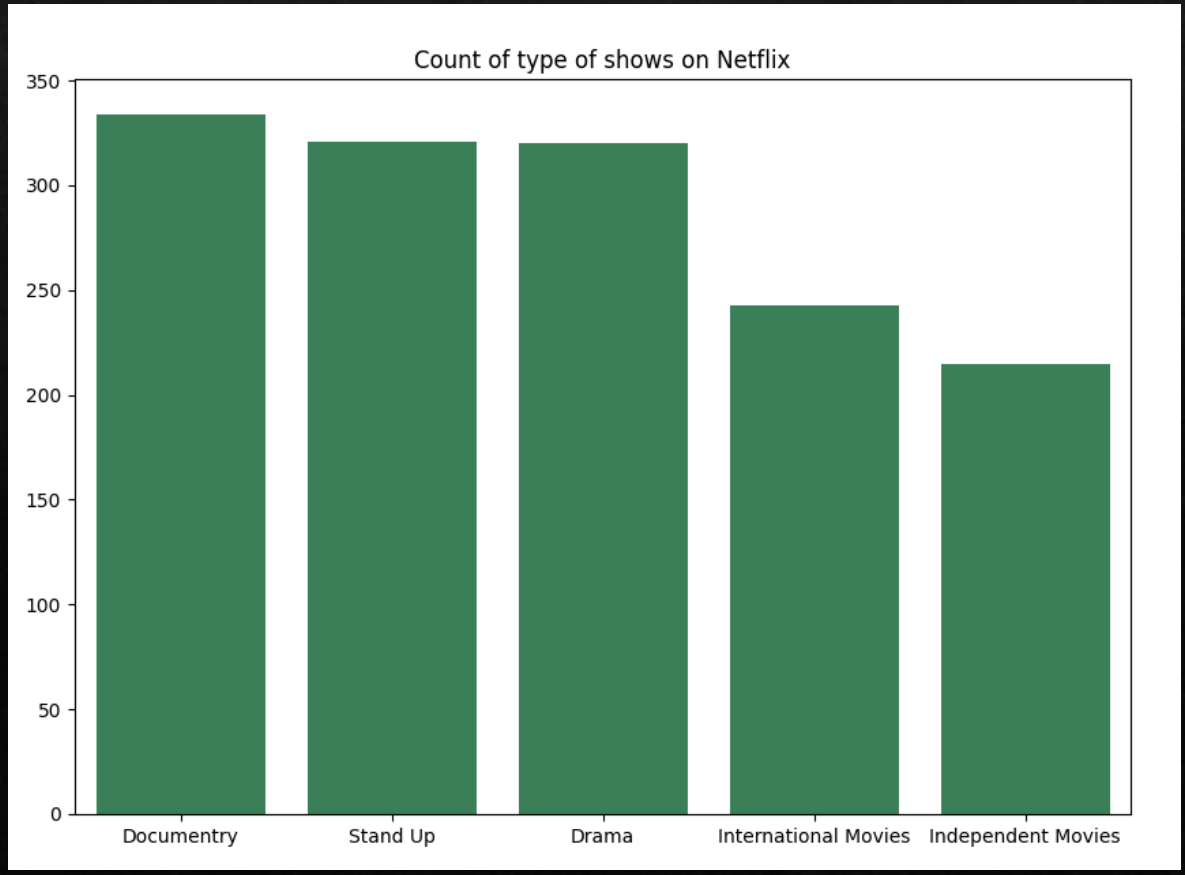


- Determining the count of different types of shows on Netflix, categorizing content into distinct types to understand the composition and variety available on the platform.

```
print(df["Type"].value_counts()[:5])
```

```
Type
Documentaries                                334
Stand-Up Comedy                              321
Dramas, International Movies                 320
Comedies, Dramas, International Movies       243
Dramas, Independent Movies, International Movies    215
Name: count, dtype: int64
```

```
plt.figure(figsize=(10,7))
plt.title("Count of type of shows on Netflix")
type= df["Type"].value_counts().head()
label = ["Documentry", "Stand Up", "Drama",
"International Movies", "Independent Movies"]
sns.barplot(x = label, y= type.values , color="seagreen")
plt.show()
```



Conclusion

- In conclusion, our analysis of the Netflix dataset has revealed valuable insights.
- The year 2019 emerged as a prolific period for content releases, particularly in movies.
- The top countries like United States and India contributing to Netflix shows and movies were identified, shedding light on the platform's global content distribution.
- Additionally, the top directors, such as Raúl Campos and Jan Suter, were recognized for their notable contributions.
- Furthermore, the content ratings, including TV-MA and TV-14, were clarified to provide a better understanding of age appropriateness.
- Documentary content emerges as the most popular genre among Netflix shows

These findings collectively contribute to a comprehensive understanding of Netflix's content landscape, aiding both content creators and viewers alike.

Findings

The findings from the analysis and conclusions highlight key aspects of Netflix content:

- Prolific Year: The year 2019 witnessed a substantial number of content releases, particularly in movies.
- Global Contribution: The United States leads in contributing content to Netflix, showcasing a diverse and globally distributed library.
- Director Contributions: Directors like Raúl Campos and Jan Suter were identified as prominent contributors to the platform.
- Content Ratings Clarification: The clarification of content ratings, including TV-MA and TV-14, enhances understanding regarding age-appropriate categorization.
- Documentary Dominance: Documentaries emerged as the most popular type of Netflix shows, indicating a strong audience preference for factual and informative content.

These findings collectively contribute to a nuanced understanding of Netflix's content landscape, providing insights valuable for content creators, platform strategists, and viewers.

Solution

- Prize Drop for More Excitement: To make things more fun, Netflix can have prize drops where you might win something cool. This can happen at different times, making watching shows even more exciting!
- Special Offers for Other Months: Netflix can give special offers and discounts during certain months. This way, you get great deals and might discover awesome shows in months you didn't expect.
- Boosting Other Directors: Not just a few, but all directors should get a chance to shine! Netflix can rank and promote shows from different directors so that everyone gets attention. This way, you might discover new favorites.
- Enhancing Independent and International Movies: Netflix can make independent and international movies even better. They might add more cool ones or make them easier to find, so you don't miss out on amazing stories from around the world.
- Netflix Variety Packs: Netflix can create special 'variety packs' where you get a mix of different types of shows bundled together. This way, subscribers can explore various genres and discover new favorites, making the Netflix experience even more enjoyable and diverse.

Thank you for dedicating your valuable time to our analysis.