Netflix Data Analysis using python

by

Mugilan K S

Description of NETFLIX Dataset

This dataset dives into Netflix's content collection, presenting details profiles for each movie and TV show. Titles, genres (Docuseries, Drama, etc.), and release years are all included. Additional details like directors, cast, durations, ratings, and descriptions might also be available. By analyzing these elements, we can explore trends in content acquisition, viewer preferences based on genre popularity and ratings, and even identify hidden gems.

Why Analyze This Dataset?

1. **Understand Viewers:** Uncover what kind of content resonates with Netflix audiences, helping tailor future offerings.

2. **Spot Trends:** Identify growing genres, director popularity, and content production patterns to inform strategies.

3. **Boost Engagement:** Learn how to keep viewers hooked by analyzing what content keeps them coming back for more.

Importing Libraries

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

Importing the Dataset

```python
df= pd.read_csv("netflix_titles.csv")
df
```

	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 Mabalane, 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, Nabi…	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	Jailbirds 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, Alam 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…

Data Exploration:

```
df.shape
```

```
(8807, 12)
```

```
df.size
```

```
105684
```

```
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 Mabalane, 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, Nabi...	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	Jailbirds 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, Alam 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...

```
df.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
     dtype='object')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      6173 non-null   object
 4   cast          7982 non-null   object
 5   country       7976 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64
 8   rating        8803 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```
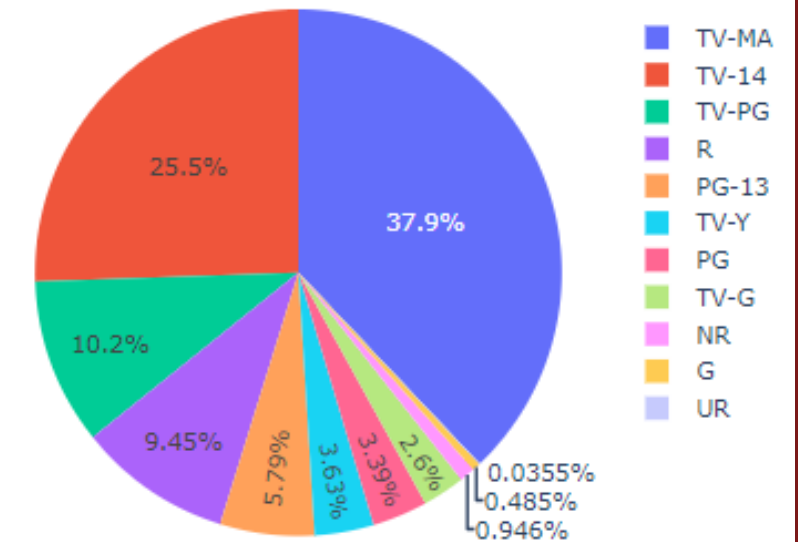
Data Visualization :

```
x= df.groupby('rating')['rating'].count().reset_index(name='counts')
x.drop(x[x['rating'].isin(['66 min', '74 min', '84 min'])].index, inplace=True)
```

```python
import plotly.express as px

pieChart = px.pie(x, values='counts', names='rating',
                  title='Distribution of content ratings on Netflix',
                  height=450, width=500)
pieChart.show()
```

Netflix offers a variety of content across different maturity levels with **[TV-MA**], **[TV-14]** being most prominent, suggesting a focus on mature audiences. Despite the emphasis on mature content, Netflix still offers a variety of content with ratings like **[PG], [G]**, and **[NR].**
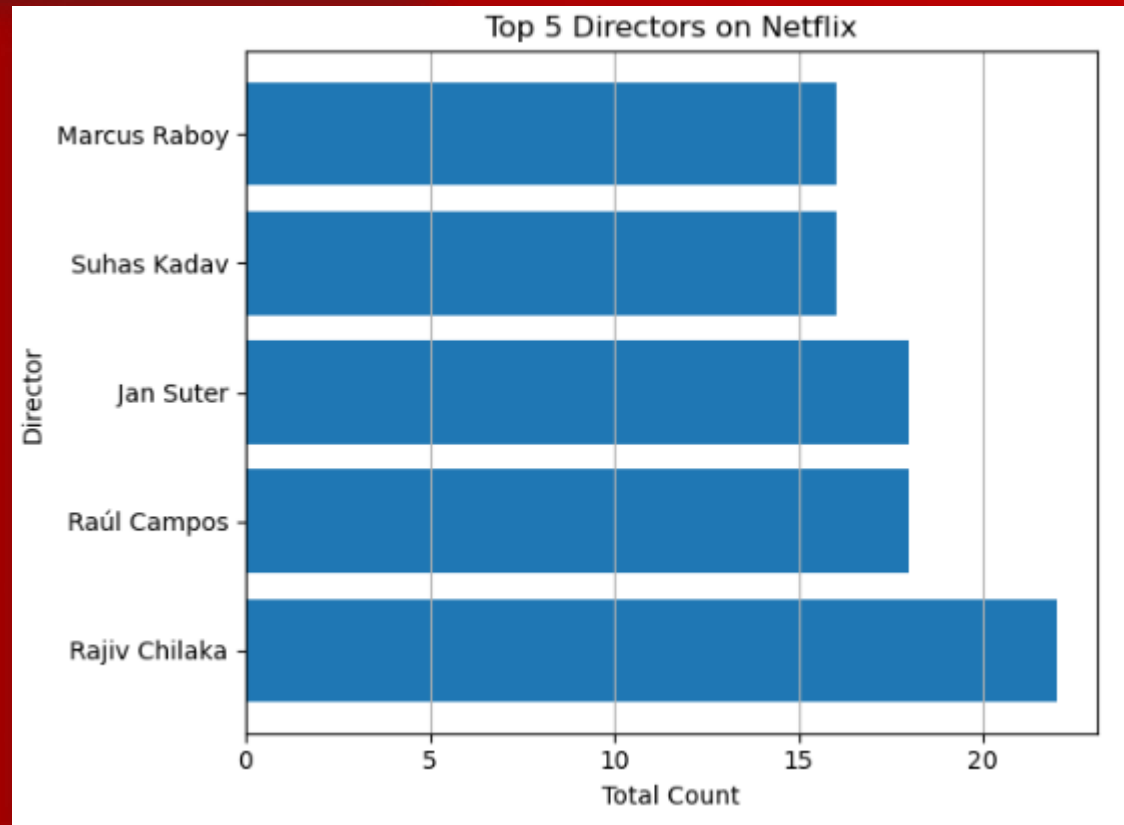

Distribution of content ratings on Netflix

Top 5 Director on NETFLIX

```python
directors_list = pd.DataFrame()
directors_list = df['director'].str.split(',', expand=True).stack()
directors_list = directors_list.to_frame()
directors_list.columns = ['Director']
directors = directors_list.groupby(['Director']).size().reset_index(name='Total Count')
directors.drop(directors[directors['Director'].isin(['Director not specified'])].index, inplace=True)
directors = directors.sort_values(by=['Total Count'], ascending = False)
top5Directors = directors.head()
```

```python
Directors = top5Directors['Director'].tolist()
Total_count = top5Directors['Total Count'].tolist()
plt.barh(Directors, Total_count)
plt.xlabel("Total Count")
plt.ylabel("Director")
plt.title("Top 5 Directors on Netflix")
plt.yticks(rotation=0)
plt.grid(axis='x')
plt.tight_layout()  )
plt.show()
```

This chart showcases the top 5 directors with the highest total count of content directed for Netflix. Rajiv Chilaka leading the pack by directing 22 movies. The following directors, Raul Campos, Jan Suter, and Suhas Kadhav, are also top director in Netflix.
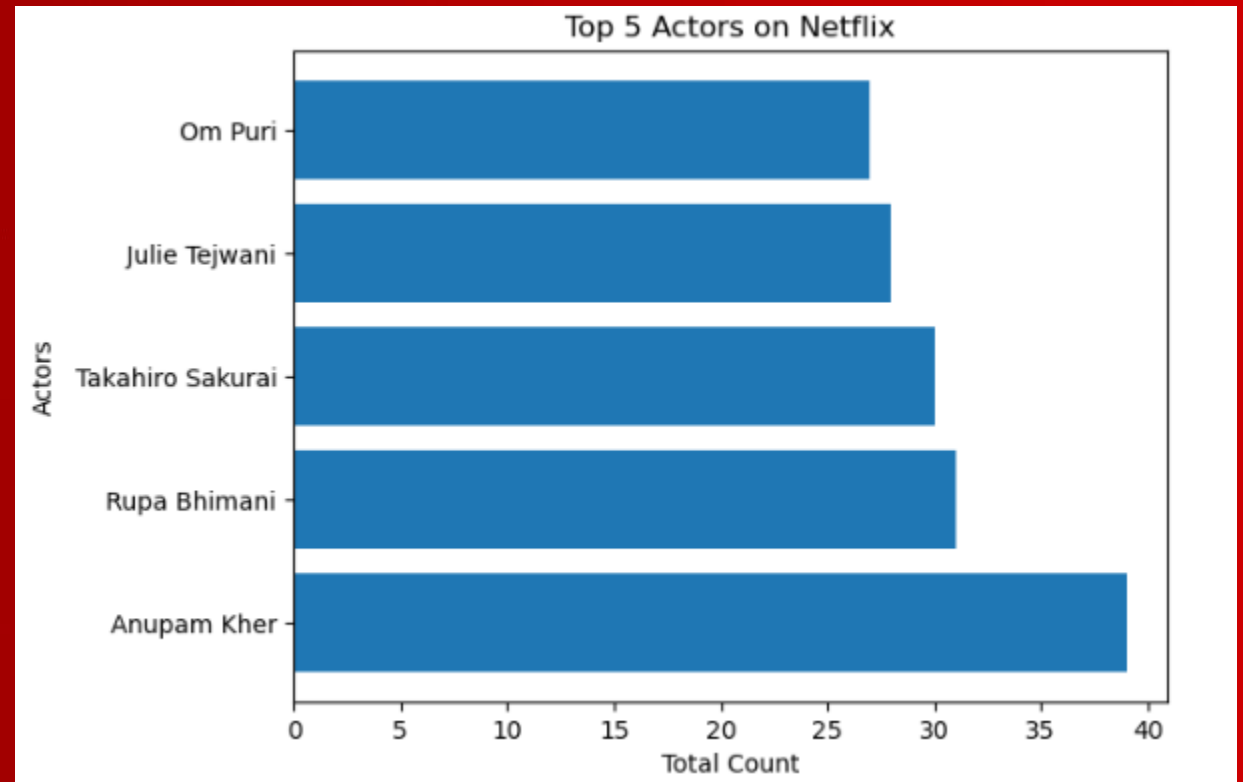
Top 5 Actor on NETFLIX

```python
cast_list=pd.DataFrame()
cast_list = df['cast'].str.split(',', expand=True).stack()
cast_list.columns= ['Actor']
actors = cast_list.groupby(['Actor']).size().reset_index(name ='Total Count')
Top5Actors = actors.head()
```

```python
Actors = Top5Actors['Actor'].tolist()
Total_count = Top5Actors['Total Count'].tolist()

plt.barh(Actors,Total_count)

plt.xlabel('Total Count')
plt.ylabel('Actors')
plt.title('Top 5 Actors on Netflix')
plt.show()
```


This chart showcases the top 5 actors with the highest number of appearances in Netflix content. Anupam Kher is the actor with the most appearances on Netflix, starring in 39 movies. The following actors, Rupa Bhomani , Takahiro Sakurai, and Om Puri, are also acted in most movies in Netflix.
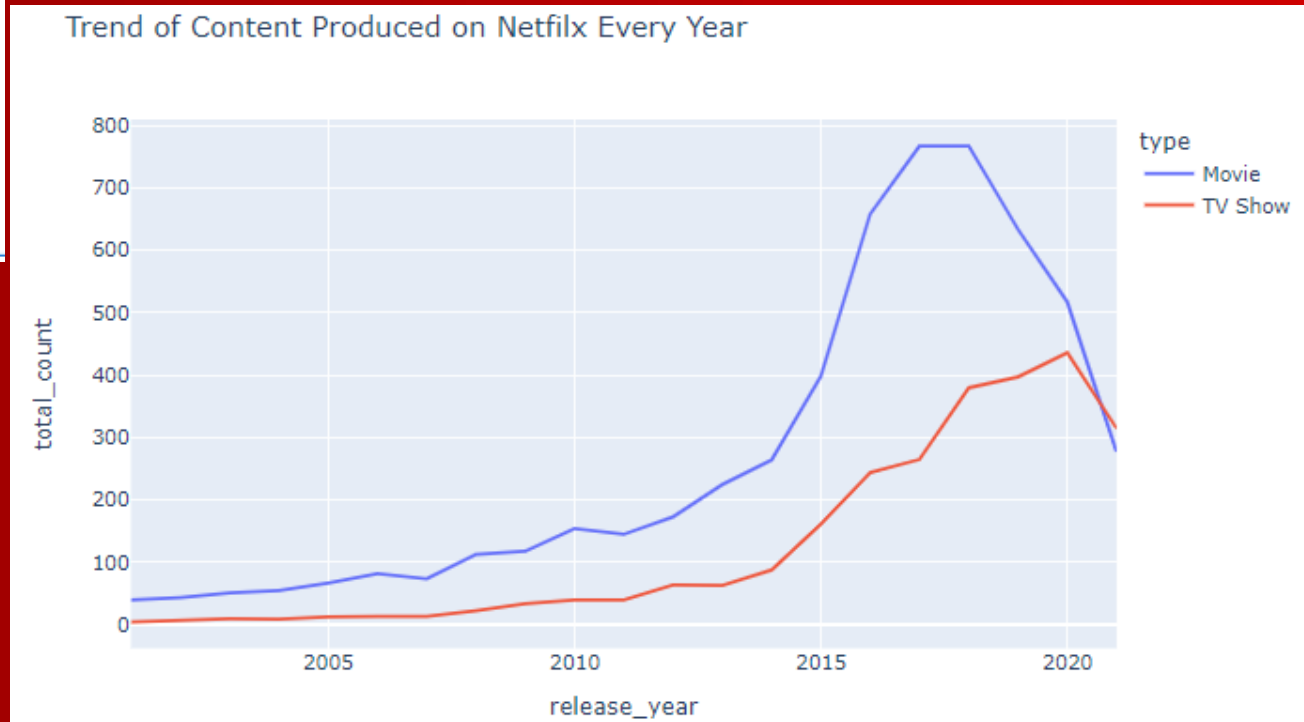
Trend of Content Produced on Netfilx Every Year

```python
df1 = df[['release_year','type']]
df2 = df1.groupby(['release_year','type']).size().reset_index(name='total_count')
release_year=df2[df2['release_year']>2000]
release_year
```

```python
import plotly.express as px

graph = px.line(release_year, x = "release_year", y="total_count", color = "type",
                title = "Trend of Content Produced on Netfilx Every Year")
graph.update_layout(
    width=800,   # Adjust width as desired (in pixels)
    height=500,  # Adjust height as desired (in pixels)
)
graph.show()
```

This graph shows an upward trend in the number of movies and TV shows produced by Netflix over the years. There's been a steady rise in the amount of content Netflix creates each year. This growth might be due to factors like subscriber base increase.
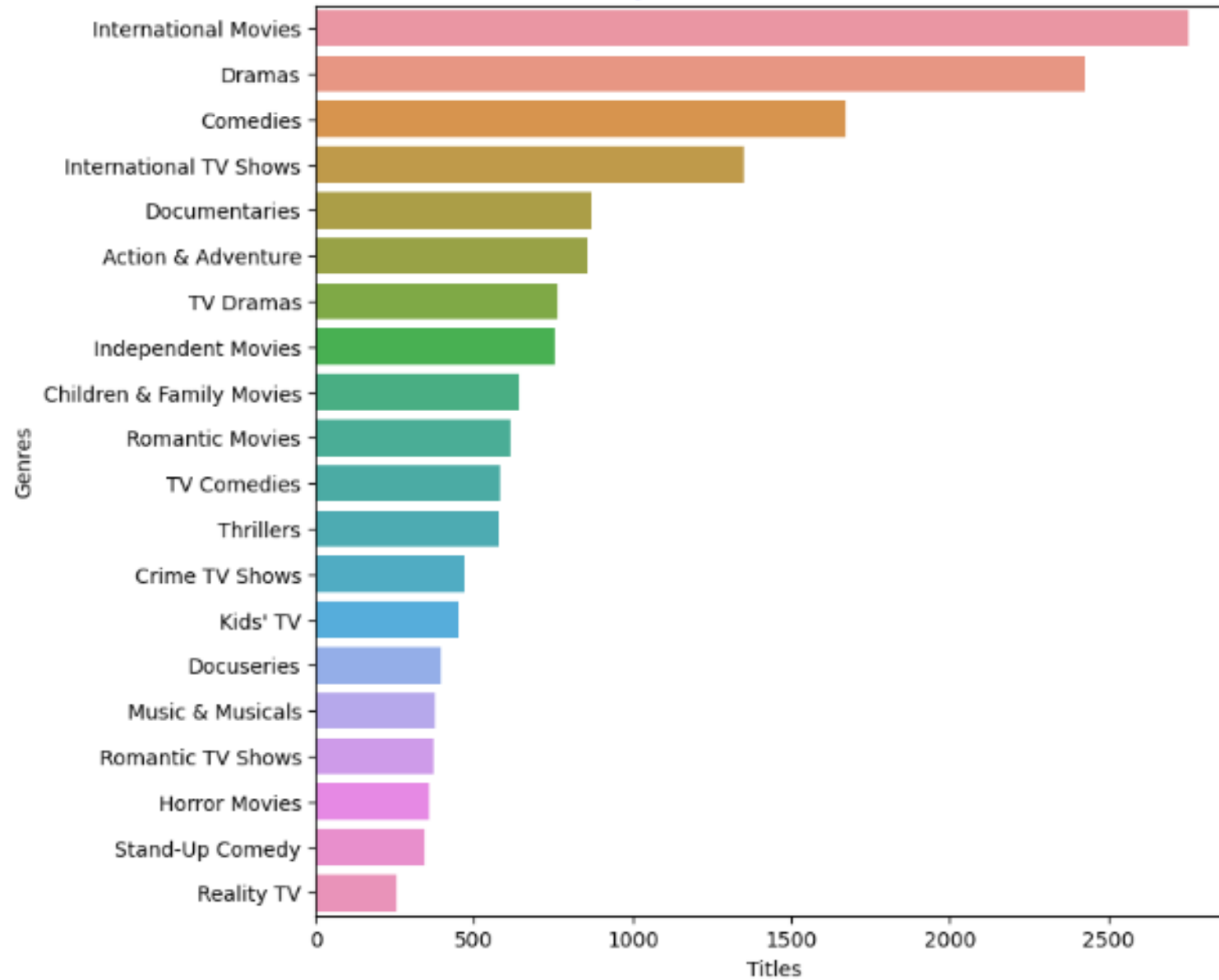
Top 20 Genres on Netflix

```python
filtered_genres = df.set_index('title').listed_in.str.split(', ', expand=True).stack().reset_index(level=1, drop=True);
plt.figure(figsize=(10,10))
custom_palette = ['skyblue', 'lightgreen', 'coral']
g = sns.countplot(y = filtered_genres, order=filtered_genres.value_counts(). index[:20])
plt.title('Top 20 Genres on Netflix',color='blue',weight='bold')
plt.xlabel('Titles')
plt.ylabel('Genres')
plt.show()
```

This chart reveals the most popular genres on Netflix, showing the number of movies and TV shows in each genre. **International Movies** and **Dramas** are among the most well-represented genres on Netflix, with a significant number of movies. Several other genres like **Comedies, Action&Adventure, Documentaries** are also have a considerable presence, indicating that Netflix offers a diverse range of content to cater to various preferences.

Top 20 Genres on Netflix

CONCLUSION

Our analysis of Netflix data reveals some key trends. We found a diverse content with prominent **[TV-MA**], **[TV-14]** content. **Rajiv Chilaka** has directed most no. of movie, while Anupam Kher has appeared in most no. of movies and leads the list of actor. Interestingly, Netflix content production has been steadily increasing both movies and series. Finally, the platform offers a wide range of genres, with **Dramas** and **International Movies** being the most popular Genre. These insights suggest that Netflix to a broad audience with a focus on original content creation.