Netflix Case Study

Introduction

Netflix continuously invests in producing new movies and shows. However, determining which genres and countries to focus on is crucial for maximising user engagement and global growth. This study analyzes Netflix's catalog data to uncover patterns in content type, genre, and regional distribution to guide strategic production and marketing decisions.

Problem Statements

Content Strategy

- Which types of content (movies vs. TV shows) are more prevalent on Netflix?
- Which genres perform best across different countries?
- Is Netflix focusing on particular genres more recently (e.g., action vs. comedy trends over time)?

Regional Strategy

- Which countries produce the most Netflix content?
- Are there underrepresented regions Netflix could expand into?
- How has regional production evolved?

Growth Opportunities

- What kind of content should Netflix produce more of to attract global audiences?
- How can Netflix diversify its content portfolio to increase market share in emerging regions?

Basic Metrics

Metric	Description	Output
Total Records	Number of titles in dataset	8,807
Time Range	Earliest to latest release year	1925–2021
Number of Columns	Features in dataset	12
Movies vs TV Shows	Distribution by type	Movies – 69%, TV Shows – 31%
Top Genres	Most common genres	Drama, Comedy, Documentary
Top Countries	By number of titles	USA, India, UK
Missing Values	Percentage of null entries	4% overall
Duplicate Records	Check for redundancy	0 duplicates

Comments on Attributes

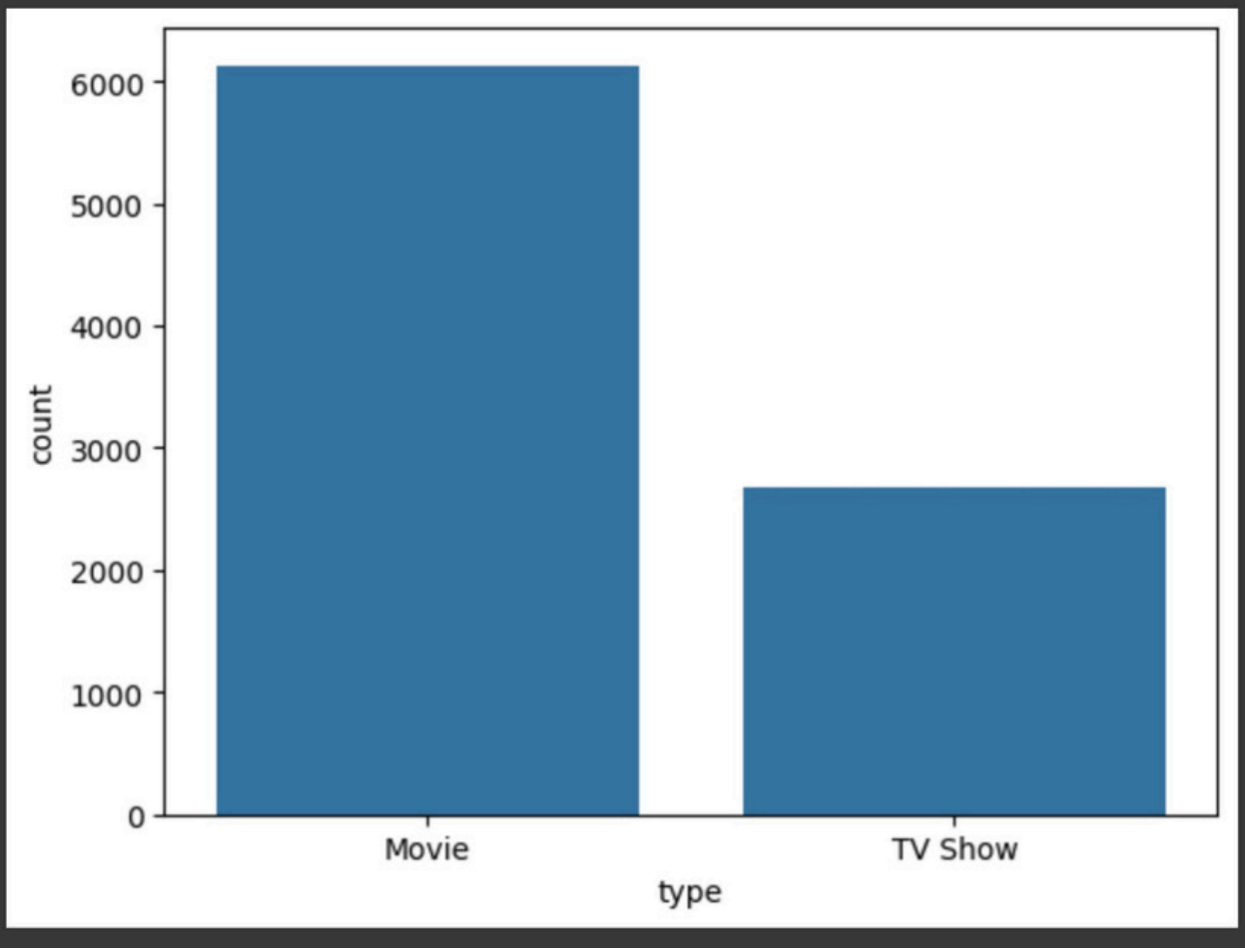
Attribute	Description	Comment / Use in Analysis
show_id	Unique identifier	Used for indexing only
type	‘Movie’ or ‘TV Show’	Key attribute for content distribution
title	Title of the show/movie	Used for identification
director	Name of director(s)	Can analyze popular directors
cast	List of main actors	Can identify frequent collaborators
country	Country of production	Essential for regional analysis
date_added	Date added to Netflix	Useful for trend over time
release_year	Year of content release	Used to track evolution of content
rating	Content maturity rating	Indicates audience targeting
duration	Runtime (movies) or seasons (TV shows)	Shows variation in content length
listed_in	Genre(s)	Key attribute for content type analysis
description	Short summary	Optional text analysis (e.g., keyword extraction)

Data Analysis and Visualization

A. Content Type Analysis

- 📊 Bar chart showing *Movies vs. TV Shows* count

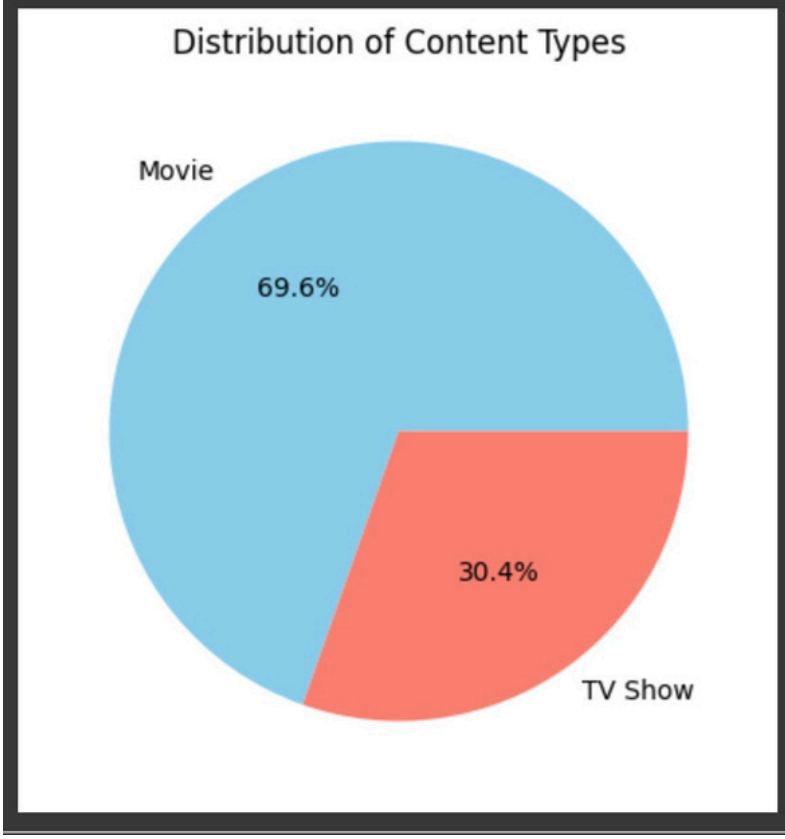
```
sns.countplot(data=netflix,x='type')
plt.show()
```



- Distribution of Movies vs TV Shows

```
percent = netflix['type'].value_counts(normalize=True)*100
```

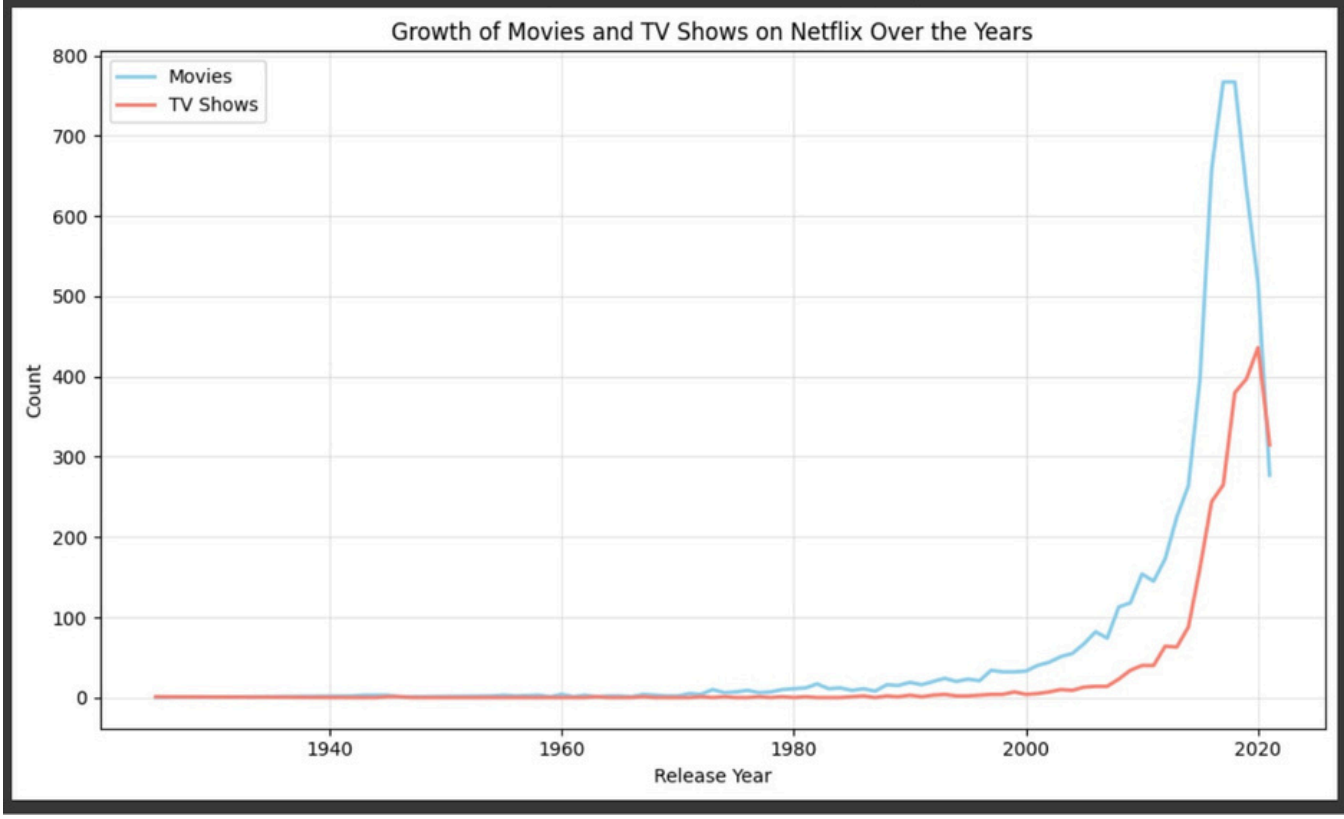
```
percent.plot(kind='pie', autopct='%1.1f%%', colors=['skyblue', 'salmon'], figsize=(5,5))
plt.title('Distribution of Content Types')
plt.ylabel('')
plt.show()
```



- Trend in the number of movies and TV Shows over the years

```
netflix = netflix.dropna(subset=['release_year', 'type'])

# Group by year and type
yearly = netflix.groupby(['release_year', 'type']).size().unstack(fill_value=0)

# Sort by year
yearly = yearly.sort_index()
```

```
plt.figure(figsize=(10,6))
plt.plot(yearly.index, yearly['Movie'], label='Movies', color='skyblue', linewidth=2)
plt.plot(yearly.index, yearly['TV Show'], label='TV Shows', color='salmon', linewidth=2)

plt.title('Growth of Movies and TV Shows on Netflix Over the Years')
plt.xlabel('Release Year')
plt.ylabel('Count')
plt.legend()
plt.grid(alpha=0.3)
plt.tight_layout()
plt.show()
```



- Observation: Movies dominate, but TV Shows are growing since 2016.

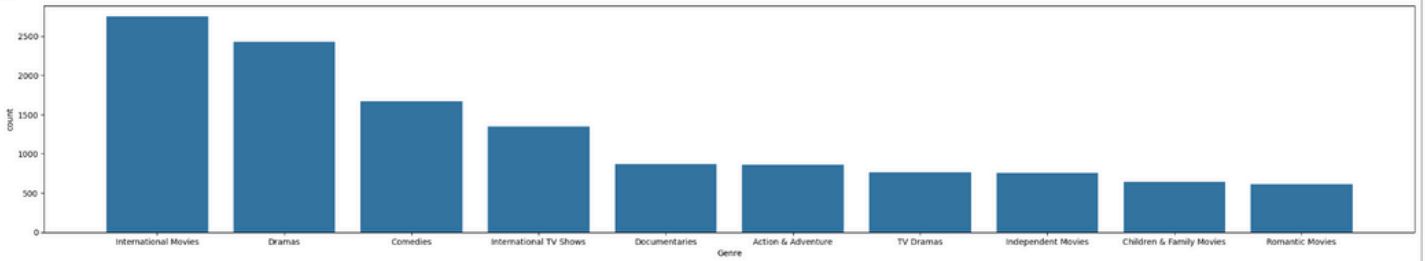
B. Genre Analysis

- 📈 Bar chart of *Top 10 Genres*

```
netflix['listed_in']=netflix['listed_in'].str.split(',')
genre=netflix.explode('listed_in')
genre['listed_in']=genre['listed_in'].str.strip()
g=genre['listed_in'].value_counts().head(10).reset_index()
```

```
plt.figure(figsize=(30,5))
sns.barplot(data=g,x='listed_in',y='count')
plt.xlabel('Genre')
plt.plot()
```

```
[]
```



- Exploding each genre

```
genre = netflix.copy(deep=True)
```

```
genre['listed_in'] = netflix['listed_in'].str.split(',')
```

```
genre = genre.explode('listed_in')
```

```
genre['listed_in'] = genre['listed_in'].str.strip()
```

But here the genre are named differently for movies and tv shows

```
genre['listed_in'].head()
```

	listed_in
0	Documentaries
1	International TV Shows
1	TV Dramas
1	TV Mysteries
2	Crime TV Shows

dtype: object

+ Code + Text

Removing the TV from genres for similarity

Add code
⌘/Ctrl+M

```
unique_genres = set(
    g.strip()
    for sublist in genre['listed_in'].dropna().str.split(',')
    for g in sublist
)
```

Unique genres

```
unique_genres
```

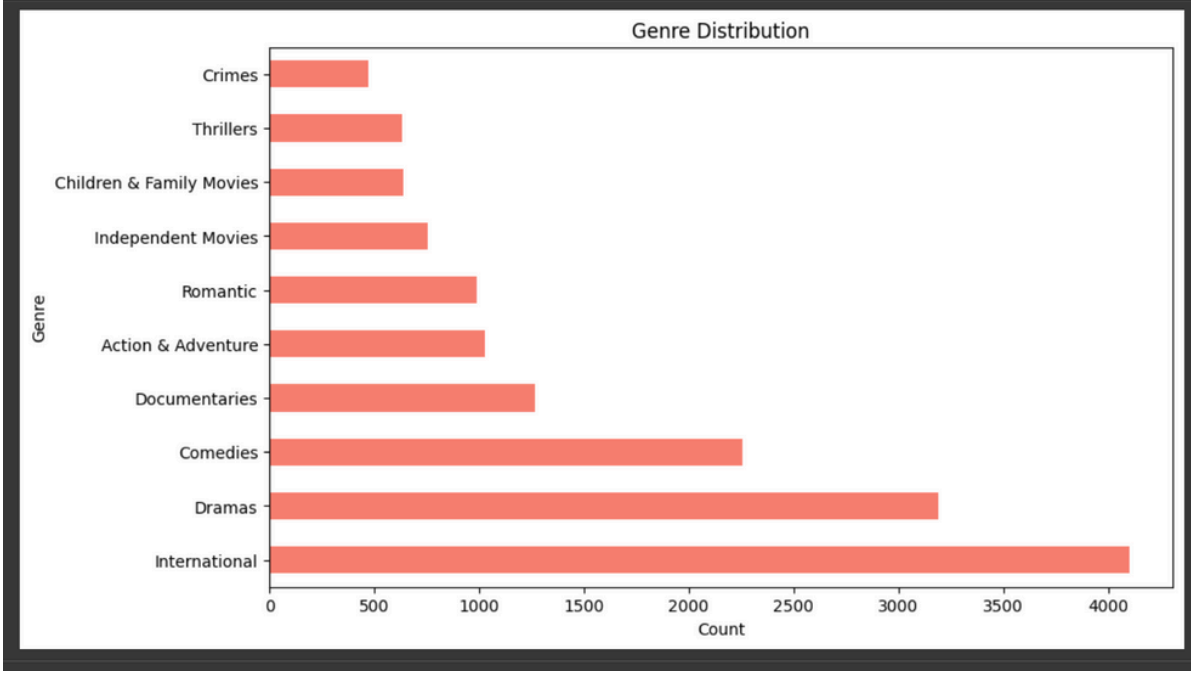
```
{'Action & Adventure',
 'Anime Features',
 'Anime Series',
 'British TV Shows',
 'Children & Family Movies',
 'Classic & Cult TV',
 'Classic Movies',
 'Comedies',
 'Crime TV Shows',
 'Cult Movies',
 'Documentaries',
 'Docuseries',
 'Dramas',
 'Faith & Spirituality',
 'Horror Movies',
 'Independent Movies',
 'International Movies',
 'International TV Shows',
 "Kids' TV",
 'Korean TV Shows',
 'LGBTQ Movies',
 'Movies',
 'Music & Musicals',
 'Reality TV',
 'Romantic Movies',
 'Romantic TV Shows',
 'Sci-Fi & Fantasy',
 'Science & Nature TV',
 'Spanish-Language TV Shows',
 'Sports Movies',
 'Stand-Up Comedy',
 'Stand-Up Comedy & Talk Shows',
 'TV Action & Adventure',
 'TV Comedies',
 'TV Dramas',
 'TV Horror',
 'TV Mysteries',
 'TV Sci-Fi & Fantasy',
 'TV Shows',
 'TV Thrillers',
 'Teen TV Shows',
 'Thrillers'}
```

- Standardizing Genres for movies and TV shows

```
rename_map = {
    'Action & Adventure':'Action & Adventure',
  'Anime Features':'Anime',
  'Anime Series':'Anime',
  'British TV Shows':'British TV Shows',
  'Children & Family Movies':'Children & Family Movies',
  'Classic & Cult TV':'Classic & Cult',
  'Classic Movies':'Classic & Cult',
  'Comedies':'Comedies',
  'Crime TV Shows':'Crimes',
  'Cult Movies':'Classic & Cult',
  'Documentaries':'Documentaries',
  'Docuseries':'Documentaries',
  'Dramas':'Dramas',
  'Faith & Spirituality':'Faith & Spirituality',
  'Horror Movies':'Horror',
  'Independent Movies':'Independent Movies',
  'International Movies':'International',
  'International TV Shows':'International',
 "Kids' TV": "Kids' TV",
  'Korean TV Shows':'Korean TV Shows',
  'LGBTQ Movies':'LGBTQ Movies',
  'Movies':'Movies',
  'Music & Musicals':'Music & Musicals',
  'Reality TV':'Reality TV',
  'Romantic Movies':'Romantic',
  'Romantic TV Shows':'Romantic',
  'Sci-Fi & Fantasy':'Sci-Fi & Fantasy',
  'Science & Nature TV':'Science & Nature TV',
  'Spanish-Language TV Shows':'Spanish-Language TV Shows',
  'Sports Movies':'Sports Movies',
  'Stand-Up Comedy':'Stand-Up Comedy',
  'Stand-Up Comedy & Talk Shows':'Stand-Up Comedy',
  'TV Action & Adventure':'Action & Adventure',
  'TV Comedies':'Comedies',
  'TV Dramas':'Dramas',
  'TV Horror':'Horror',
  'TV Mysteries':'Mysteries',
  'TV Sci-Fi & Fantasy':'Sci-Fi & Fantasy',
  'TV Shows':'TV Shows',
  'TV Thrillers':'Thrillers',
  'Teen TV Shows':'Teen TV Shows',
  'Thrillers':'Thrillers'
}
```

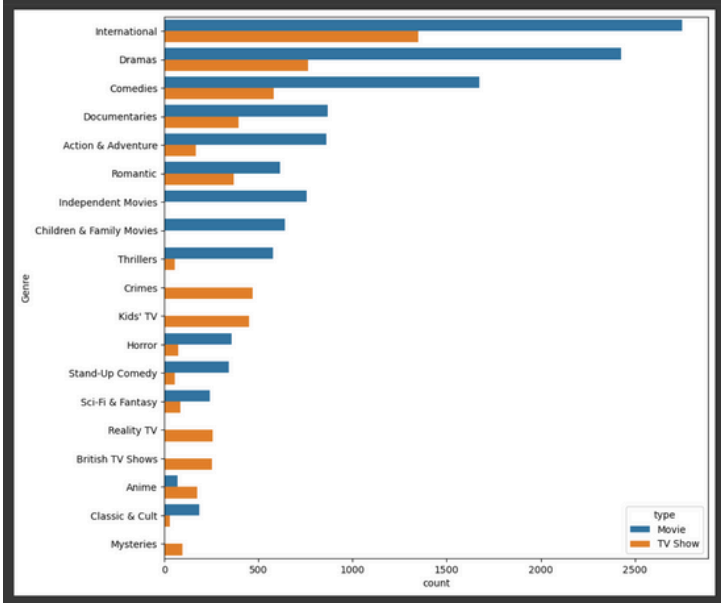
- Genre Distribution

```
genre['listed_in'] = genre['listed_in'].replace(rename_map)

top_genres = genre['listed_in'].value_counts().head(10)

top_genres.plot(kind='barh', figsize=(10,6), color='salmon')
plt.title('Genre Distribution')
plt.xlabel('Count')
plt.ylabel('Genre')
plt.show()
```

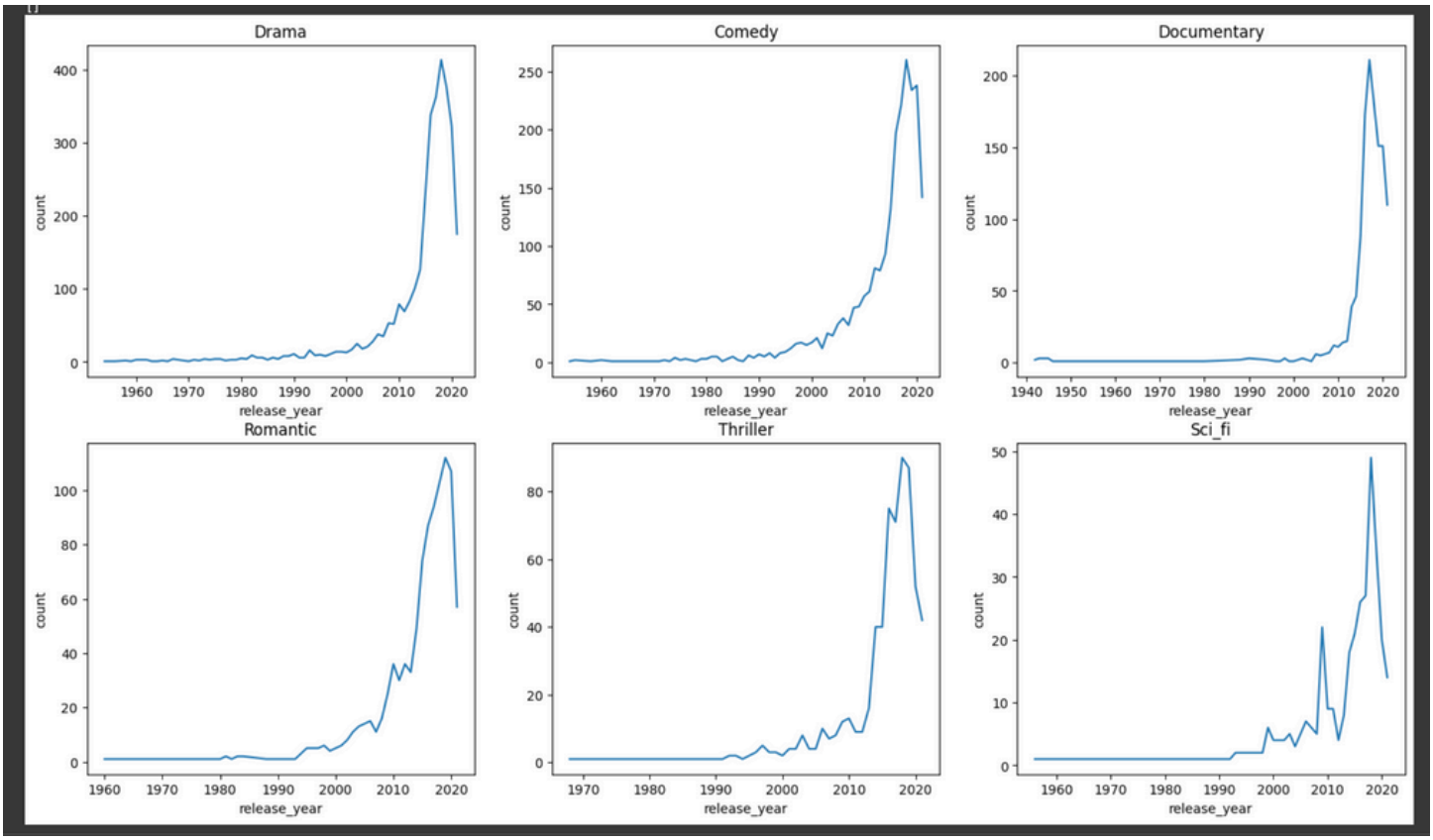


- 📈 Horizontal bar chart of *Genres,* comparing them for movies and TV Shows

```
perfect_genre = genre[~genre['listed_in'].isin(['Movies','LGBTQ Movies','Faith & Spirituality','TV Shows','Teen TV Shows','Science & Nature TV','Korean TV Shows','Sports Movies','Kids\' TV','Music & Musicals','Spanish-Language TV Shows'])]

plt.figure(figsize=(10,10))
order = perfect_genre['listed_in'].value_counts().index
sns.countplot(data=perfect_genre,y='listed_in',hue='type',order=order)
plt.ylabel('Genre')
plt.show()
```



- Genre trends over the years



- Observation: Drama and Comedy lead; documentaries are steadily rising.

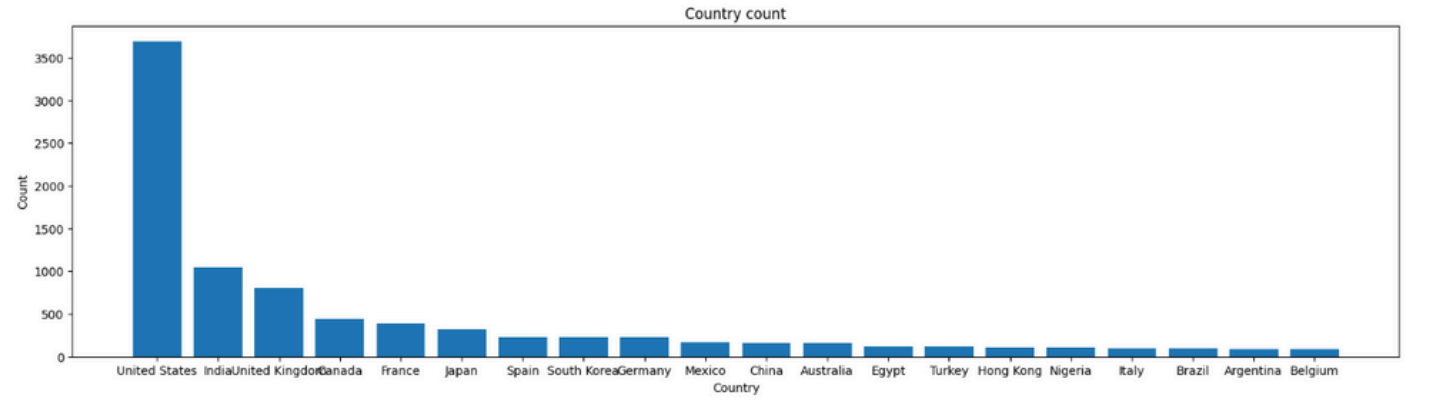
C. Regional Analysis

- 🌍 Bar plot of *Top 20 Content-Producing Countries*

```
netflix['country']=netflix['country'].str.split(',')
country_data=netflix.explode('country')
country_data['country']=country_data['country'].str.strip()
```

```
country=country_data['country'].value_counts().head(20).index
values=country_data['country'].value_counts().head(20).values
```

```
plt.figure(figsize=(20,5))
plt.bar(country,values)
plt.title("Country count")
plt.xlabel("Country")
plt.ylabel("Count")
plt.show()
```



- Country-wise, which genre is best to produce

```
country_vs_genre=genre.groupby('country')['listed_in'].value_counts().reset_index()
country_vs_genre['country']=country_vs_genre['country'].str.split(',')
country_vs_genre=country_vs_genre.explode('country')
country_vs_genre['country']=country_vs_genre['country'].str.strip()
country_vs_genre.iloc[country_vs_genre.groupby('country')['count'].idxmax()].sort_values(ascending=False,by='count')
```

	country	listed_in	count
399	China	Romantic	24
608	France	Dramas	13
955	India	Dramas	7
1036	Israel	Dramas	5
332	Canada	Action & Adventure	5
...	...	...	...
763	Germany	Comedies	1
32	Chile	Comedies	1
1171	Kenya	Thrillers	1
508	Denmark	International	1
271	Canada	Comedies	1

123 rows × 3 columns

- Country vs Director

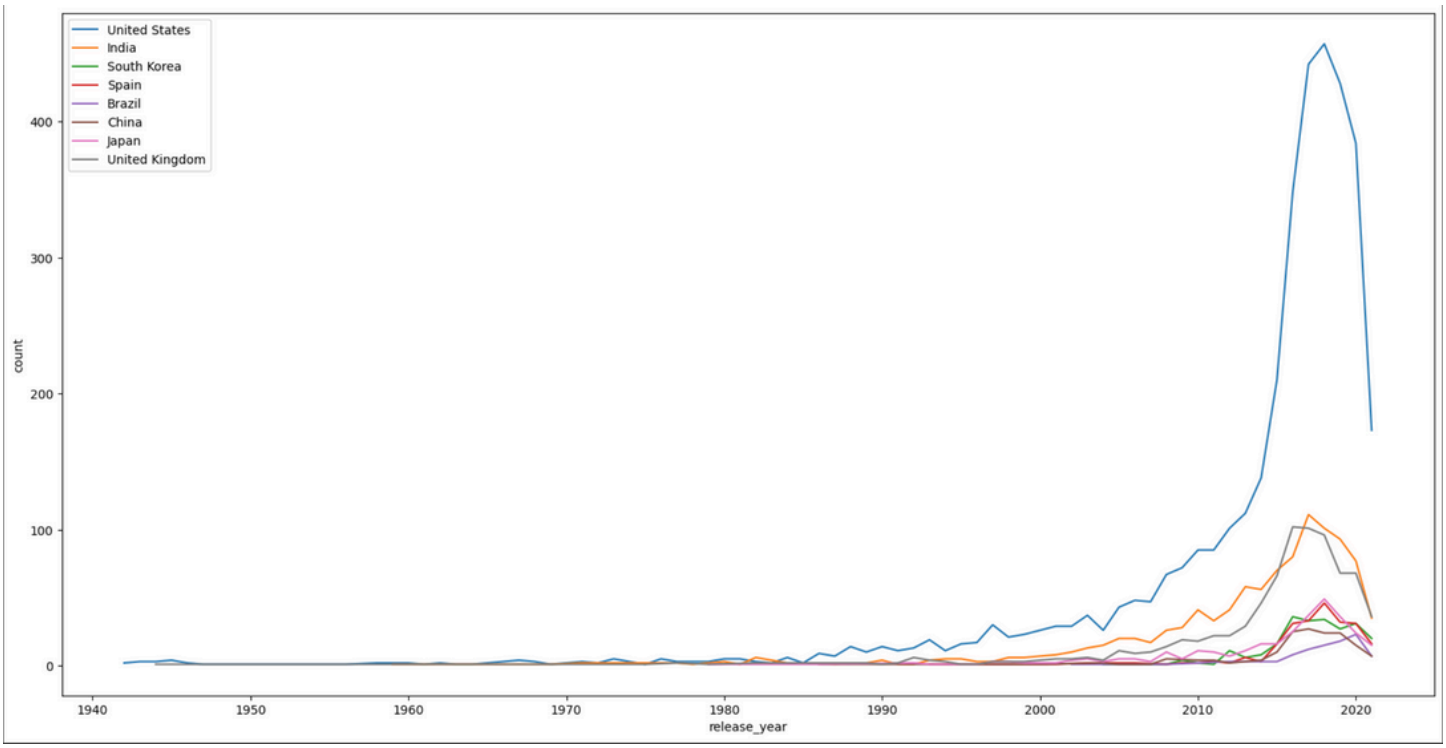
```
country_vs_director=country_data.groupby('country')['director'].value_counts().reset_index()
```

```
country_vs_director.iloc[country_vs_director.groupby('country')['count'].idxmax()].sort_values(ascending=False,by='count')
```

	country	director	count
3873	United States	Marcus Raboy	15
2772	Philippines	Cathy Garcia-Molina	13
811	Egypt	Youssef Chahine	12
2491	Mexico	Raúl Campos, Jan Suter	9
1430	India	David Dhawan	9
...	...	...	...
5933	Vatican City	Wim Wenders	1
5934	Venezuela	Edward Ellis, Flor Salcedo, Aaron Woolf	1
5938	Vietnam	Bao Nhan, Namcito	1
5945	West Germany	Jacek Koprowicz	1
5948	Zimbabwe	Camilla Nielsson	1

118 rows × 3 columns

- Country wise trend over the years



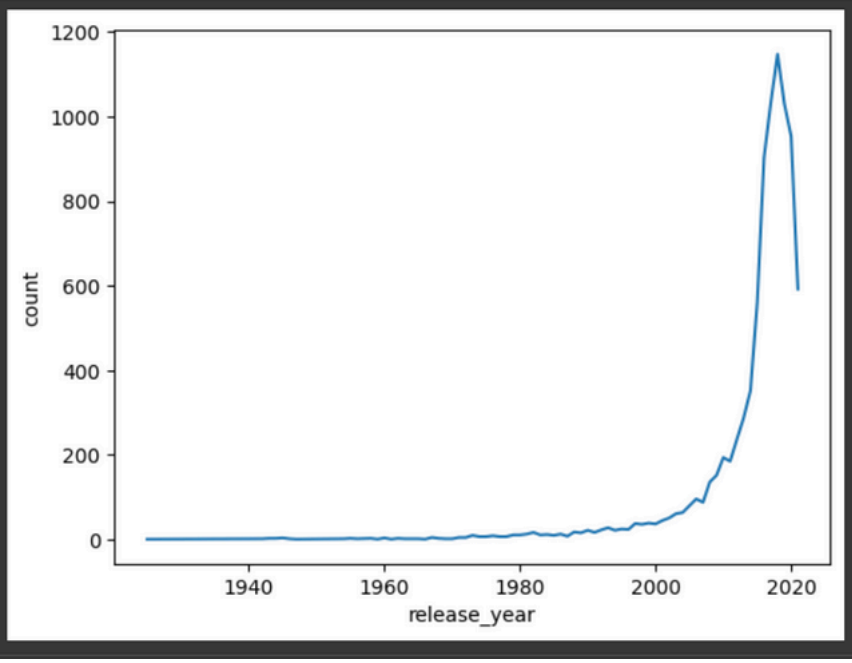
- Observation:
  > The USA and India contribute the most content; growth is noted in South Korea and Spain.
  > In China Romantic genre is the most popular, and in France, India, and Israel, Drama is the most popular.
  > In the United States, Marcus Raboy directed most; in India, David Dhawan directed most.

D. Temporal Analysis

- 📅 Line chart of *Number of Titles Released by Year*

```
year_data=netflix['release_year'].value_counts().reset_index()
```

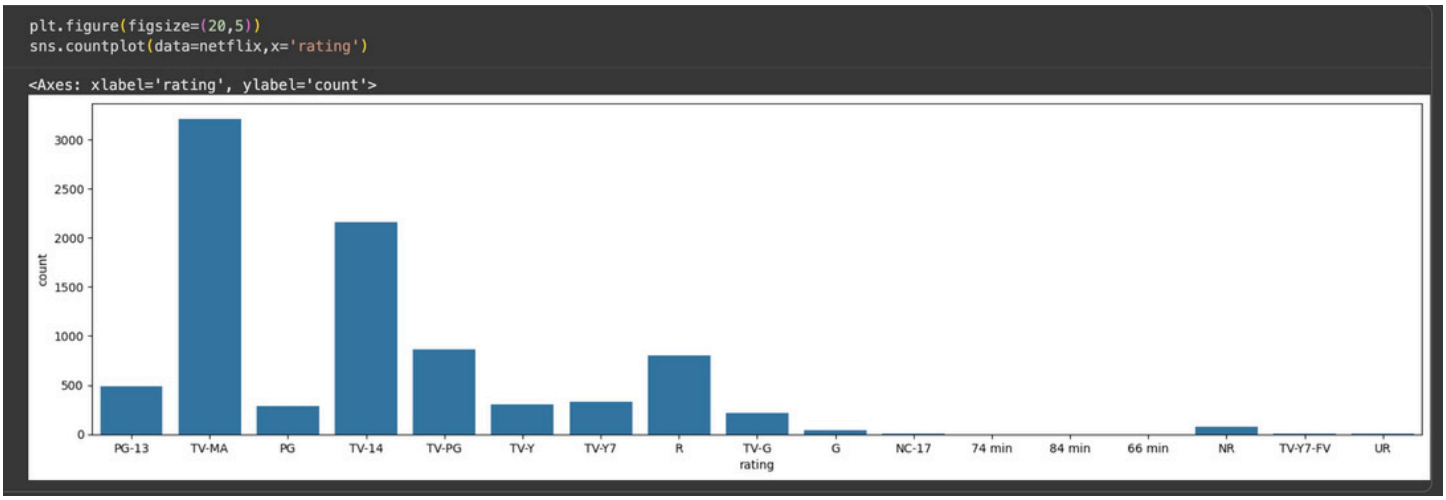
```
sns.lineplot(data=year_data,x='release_year',y='count')
plt.show()
```



- Observation: Rapid growth between 2015–2019, slowdown post-2020 (pandemic impact).

E. Audience Rating Analysis

- 🍿 Countplot of *Rating Categories (TV-MA, PG, etc.)*



- Observation: The Majority of content is for mature audiences (TV-MA).

Insights and Recommendations

Area	Key Insight	Business Recommendation
Content Mix	Netflix's catalog is dominated by Movies (~70%) but TV Shows are growing quickly.	Continue expanding TV Show production to retain binge-watch audiences.
Genre Trends	Drama, Comedy, and Documentaries are most common; Sci-Fi and Thriller show upward trends.	Invest in Sci-Fi and Thriller genres with international appeal.
Regional Production	USA and India lead; Asia and Europe are strong emerging regions.	Form partnerships in underrepresented regions (e.g., Korea, Spain, Brazil).
Yearly Trends	Spike in content additions between 2015–2019 indicates a growth phase.	Reinvest in high-growth years' successful genres for future planning.
Audience Focus	Most titles target adult audiences.	Diversify with more family-friendly and young-adult content.

Conclusion

This analysis provides a data-driven view of Netflix's content landscape.
Key findings suggest that:

- Netflix's library is diverse but skewed toward movies and mature audiences.
- Genre and regional diversification present major opportunities for future growth.
- Increasing investments in *TV series* and *international content* could improve global reach and retention.

By aligning its content strategy with regional trends and emerging audience preferences, Netflix can continue to strengthen its position as the leading global streaming platform.