📽️ 🎬 🍿 Netflix Trends Analysis

Author: Ana Pesut

🎯 Goals

This analysis will examine various characteristics of the film industry in order to monitor emerging trends and identify strategic investment opportunities that are likely to drive significant traffic to our streaming platform. In order to do so, it is crucial to follow these given requirements:

1. We are first going to import necessary libraries as tools for our analysis and load our dataset, as always.

2. We will identify the most productive years in filmmaking, which will help us explore the underlying factors contributing to success at these particular points in time. This will be visualized through the medium of **line chart**. Additionally, we will take the opportunity to present a **pie chart** illustrating the proportional shares of series and films, providing further insight into content distribution.

3. In the next step, we will analyze which movies are most popular based on their duration (or runtime). We will also examine the frequency of genres, in order to determine which genres are most favoured by audiences. An interactive graph presenting this analysis will be created in the form of **histogram**

4. We will analyize quality and popularity by comparing IMDb and TMDB ratings using **boxplots** for both platforms, distinguishing between series and films. This will help us determine which platform has stricter user ratings and whether differences exist across genres or other characteristics. Furthermore, a **scatterplot** will be used to show TMDB ratings and popularity, providing management with concrete insights into whether higher ratings reliably predict greater popularity or if notable exceptions exist. The interactivty of the graphs will be signficantly enhanced sing features such as **hover info**, which will improve accessibility and provide a refined, user-friendly final touch to our visualizations.

5. In the final step, it is required of us to provide a brief summary of the main points from our analysis, as well as offer suggestions based on these conclusions for the purpose of making informed strategic investments.

⬅️ 👩‍💻 Step 1: Importing Libraries, Loading the Dataset and Showing Sample Data

	id	title	type	description	release_year	age_certification	runtime	g
0	ts300399	Five Came Back: The Reference Films	SHOW	This collection includes 12 World War II-era p...	1945	TV-MA	51	['documenta
1	tm84618	Taxi Driver	MOVIE	A mentally unstable Vietnam War veteran works ...	1976	R	114	['drama', 'c
2	tm154986	Deliverance	MOVIE	Intent on seeing the Cahulawassee River before...	1972	R	109	['drama', 'a 'th 'euro
3	tm127384	Monty Python and the Holy Grail	MOVIE	King Arthur, accompanied by his squire, recrui...	1975	PG	91	['fantasy', 'a 'cor
4	tm120801	The Dirty Dozen	MOVIE	12 American military prisoners in World War II...	1967	NaN	150	['war', 'a

📈 🎬 Step 2: Finding the Most Productive Years and Comparing the Fields (Movies vs. Series)

Grouping Movies and Shows by Release Years
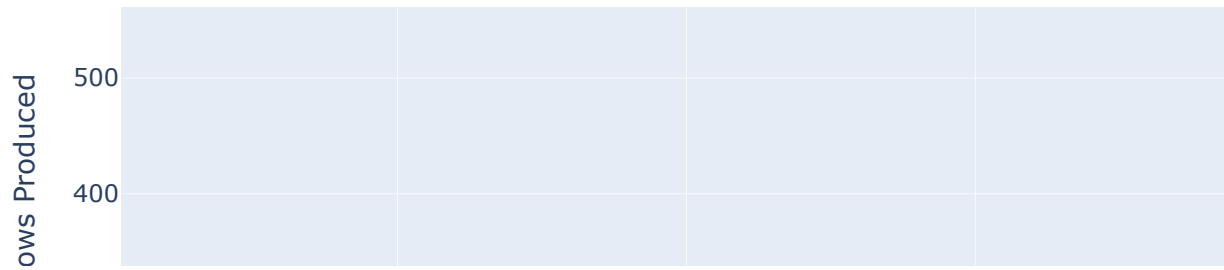
The data revealse distinct patterns in movie and show production over time. **Movie production reached** its peak in **2019, with 525 titles released**, representing the highest single-year output in the dataset. This peak was followed by a decline in subsequent years. **Television show production** exhibited a somewhat different trajectory, with both **2020 and 2021** recording identical peak production levels of **314** each. This sustained high output over consecutive years suggests a structural shift in the entertainment industry, rather than a temporary spike. Further insights and possible reasons for these trends will be provided in the main conclusion paragraph at the end of this analysis.

Distribution of Movie and SHOWS Share

In this extension of our analysis, we can see that the **movie industry** demonstrates a significant market dominance, comprising 64% of the total content portfolio, while television shows account for the remaining 36%. This distribution reveals a nearly 2:1 ratio, favoring film production over serialized content. However, we have seen in the previous graph that the movie production doesn't show consistent efforts, deviating from its usual production rate in the year 2019. Potential reasons for this shall be outlined in the final conclusion.

📊 ⏱️ Step 3: Short and Striking or Long and Lasting?

Distribution of Movies by Duration
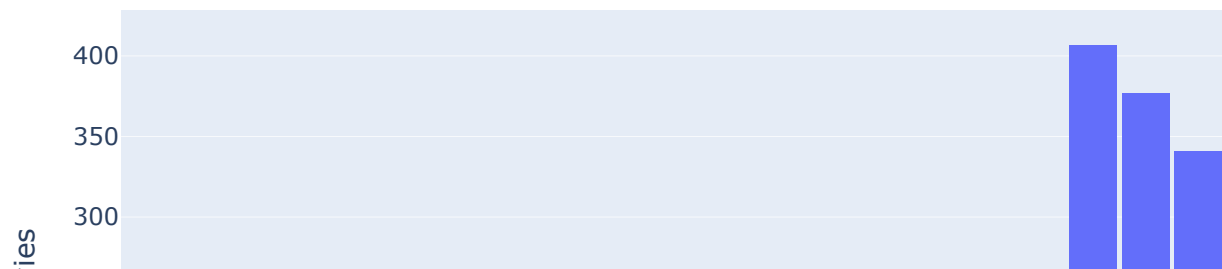
In this analysis we can see that the most preferred movies are those that belong to the mid-range duration: from **95-99 minutes, with the data slightly raising towards the 90-94 minute period**. Over **400 movies** alone belong to this timeframe, which makes up a considerable amount of about 10% of total movie content portfolio. We could count that in the following way:

```
Movies make a total of: 10.68% in the movie dataset.
```

The most popular movies fall within the **90-94 minute range, with approximately 407 titles alltogether**in this duration category. Overall,the data suggests that the ideal movie length gravitates toward around an hour and a half-likely because it allows for a fast-paced narrative with plenty of engaging content packed into a shortertime frame.

The most popular genres
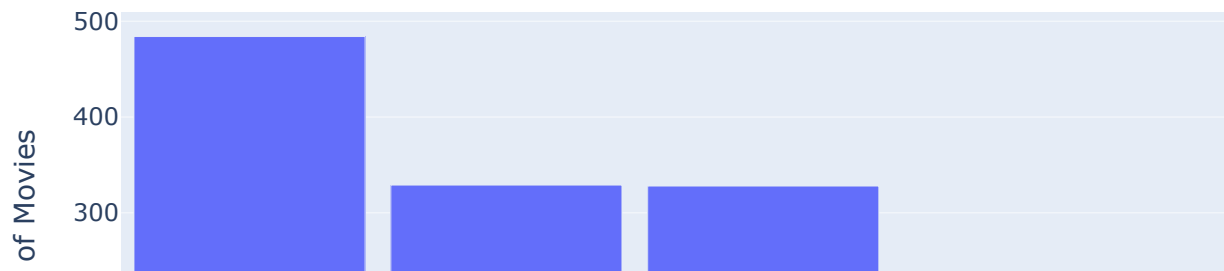
The most popular genre according to our bar chart, is **comedy**, with a total of **484 movies**.

📦 Step 4: Making Boxplots for IMDb and TMDb scores

IMDb Scores: Movies vs. Series

The boxplots show that both SHOW and MOVIE **IMDb ratings are relatively high**, but SHOW has a slightly higher median (7.1 vs. 6.4) and a more concentrated interquartile range (6.4–7.7) compared to MOVIE (5.6–7.1), suggesting more consistent ratings. SHOW also has higher minimum (2 vs. 1.6) and maximum values (9.5 vs. 9.1), indicating a slightly narrower and higher overall rating spread. Overall, **SHOW appears to have better and more consistent ratings than MOVIE**

TMDb Scores: Movies vs. Series

Once again, SHOWS seem to receive higher TMDb ratings than MOVIES. Both categories show similar consistency in ratings, as indicated by nearly identical differences between their medians and interquartile ranges. However, SHOWS tend to score higher overall, with even their lower extremes surpassing those of MOVIES. This suggests that SHOWS generally maintain a stronger baseline in audience ratings.

TMDb scores vs TMDb popularity
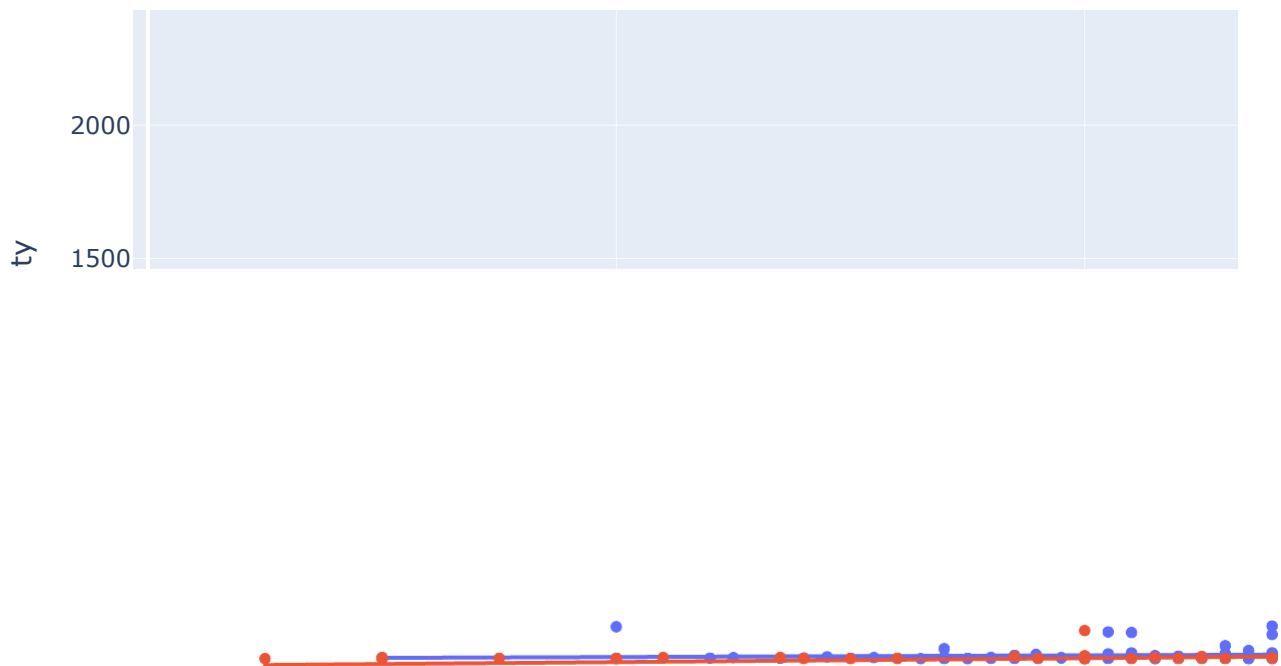
The OLS trendlines for SHOWS and MOVIES fully overlap, indicating no meaningful difference in how TMDb score relates to popularity between the two categories. Additionally, the lines appear relatively flat, suggesting that popularity does not change significantly with the TMDb score. This implies a weak or no clear relationship between ratings and popularity for both SHOWS and MOVIES in this dataset.

📌 Conclusion:

Based on the previous analysis we can give a few recommendations for a head start in streaming industry:

1. Balance Content: Keep a strong movie catalog but increase investment in high-quality TV shows, which have growing production and better audience ratings.

2. Focus on Popular Formats: Prioritize movies around 90–99 minutes and comedy genres to match viewer preferences.

3. Enhance Discoverability: Since ratings don't predict popularity well, improve personalized recommendations and targeted marketing to boost engagement.

4. Adapt to Trends: Stay aligned with the shift toward more TV show production seen in recent years to meet evolving audience demand.