

NETFLIX EDA

Below is the exploratory data analysis of the Netflix dataset, key findings, some interesting patterns, and conclusions.

Key Findings:

1. Content Distribution:

- Movies nearly 70% of the content on Netflix, while TV Shows are around 30%.
- The number of movies and TV shows has increased regularly over the years, with a visible spike between the years 2016-2019.

2. Genre Analysis:

- Dramas and Comedies are very common across both movies and TV shows.
- The percentage of the Reality and Documentary genres is higher in TV Shows compared to movies.
- Action and Thriller genres are specific to movies.

3. Country of Origin:

- The United States dominates the library.
- Increasing diversity with meaningful contributions from the United Kingdom, South Korea, and India.

4. Content Duration:

- Most movies fall within the 90-120 minute range.
- TV shows usually are just about seasons of minutes assuming each season is counted as 400 minutes.

5. Ratings Distribution:

- TV-MA and TV-14 are the rating categories that come up the most, indicating that the content is mature in nature.

6. Release Patterns:

- Releases have increased consistently over the course of the last decade.
- Fridays and Thursdays were apparently the days of preference for new content addition.
- Months with the highest volumes of new content addition were typically January and November.

7. Directors and Cast:

- The platform has directors that appear very frequently, including David Fincher and Ava DuVernay.
- There are certain genres associated with actors; for example, Adam Sandler is very tightly related to Comedy.

Some interesting trends from this data include:

1. Genre Evolution: Genres were more popular at different points in time, and some have become more frequent recently.
2. Hybrid Genres: Genres display a high degree of co-occurrence, therefore indicating hybrid genres or subgenres. For example, international movies and drama, dramas and independent movies.
3. Content Diversity: A trend of diversification of content in terms of genres and countries of origin is observed.
4. Content Age vs. Popularity: High ratings can be held for older content. Thus, classic or highly regarded content does not perish.
5. Seasonal Patterns: Addition of content seems to follow seasonal patterns, as there are increased volumes in some months of the year, such as January and November, and lesser volumes during summer months.

Conclusions and Further Analysis:

1. Their content strategy is diversified by genre and global representation, targeted at a wide, internationally spread audience.
2. The high presence of TV-MA and TV-14 rated programs suggests the platform really is into mature content.
3. The positive trend for every year infers Netflix would continue increasing the size of its content library to thrive in the streaming market.
4. Most popular genres, directors, or actors could provide insights about content acquisition and productions by Netflix.

Other possible directions:

1. Viewer Engagement: Understand how a set of viewer data mixes with features of content, such as genre, duration, or origin, to result in viewer engagement.
2. Regional Preferences: Better understand regional content preferences in order to further inform localization strategies.
3. Content Longevity: Understand what underlies the long-term popularity of older content.
4. Predictive Modeling: Model success at the front end using a variety of possible features, including genre, cast, director, and time of release.
5. Competitor Analysis: Comparison of Netflix's content strategy with other streaming platforms
6. Financial Analysis: Correlate characteristics with production cost and revenues in a way to enable an optimization of strategies in investment across content types.