Project Report: Analysis of Netflix Dataset

1. Introduction

In this project, we analyzed a Netflix dataset to gain insights into its content offerings. The dataset contained 8807 rows and 12 columns, and we aimed to explore various aspects of the content, including genres, ratings, release trends, and geographical distribution.

2. Data Cleaning

2.1 Dropping Useless Columns

The 'id' and 'description' columns were identified as useless for our analysis and were dropped from the dataset.

2.2 Handling Null Values

The dataset had null values that needed to be addressed. We used different strategies for imputing missing values based on the respective columns. For 'director,' we filled null values with 'unknown.' For 'country,' 'date_added,' 'rating,' and 'duration,' we used the mode to impute the most frequent values.

2.3 Checking for Duplicates

We checked for full duplicates in the dataset and found that there were none, ensuring data integrity.

3. Exploratory Data Analysis

3.1 Separating Countries

We separated the countries listed in the 'country' column to enable more in-depth analysis of regional content.

3.2 Correcting Wrong Values in 'rating'

We identified and replaced incorrect values in the 'rating' column with the correct ones.

3.3 Adding Time Features

We added 'month' and 'year' columns to facilitate temporal analysis.

4. Content Insights

4.1 Movies vs. TV Shows

The dataset contained approximately twice as many movies as TV shows, indicating a higher focus on movie content.

4.2 Best Months to Add Content

Based on the year 2020 data, we identified February, March, and May as the best months to add new content to Netflix.

4.3 Genre Distribution

Netflix offers a wide range of genres, precisely 20 in total. We analyzed the most common genres for movies and TV shows.

4.3.1 Most Common Movie Genres

The top three most common movie genres were:

1. International Movies (2752 movies)

2. Dramas (2427 movies)

3. Comedies (1674 movies)

4.3.2 Most Common TV Show Genres

The top three most common TV show genres were:

1. International TV Shows (774 shows)

2. TV Dramas (696 shows)

3. TV Mysteries (98 shows)

4.4 Release Trends

4.4.1 Movies Release Trends

Movies released on Netflix showed consistent numbers until 2010, after which they saw a significant increase from 2010 to 2020.

4.4.2 TV Shows Release Trends

TV shows released on Netflix followed a similar trend, with consistent numbers until 2010, followed by a substantial rise in the period 2010-2020.

4.5 Geographical Distribution

Both movies and TV shows saw the greatest number of releases in the USA, followed by India and the United Kingdom.

4.6 Duration and TV Show Seasons

The most common movie duration on Netflix fell within the range of 90-100 minutes. For TV shows, the most popular season was always Season 1.

4.7 Ratings

The majority of movies and TV shows on Netflix carried the 'TV-MA' rating, followed by 'TV-14' and 'TV-PG.'

4.8 TV Shows with Maximum Seasons

The TV show 'Grey's Anatomy' had the highest number of seasons (17), followed by 'Supernatural' and 'NCIS,' each with 15 seasons.

5. Conclusion

The analysis of the Netflix dataset provided valuable insights into the content offerings of the platform. The dataset was thoroughly cleaned to ensure accurate results. The examination of various attributes allowed us to identify trends such as release patterns, genre popularity, and regional preferences. These insights can aid Netflix in making informed decisions related to content acquisition and release strategies.

Recommendations

Based on the analysis, the following actions are recommended to Netflix:

1. Focus on expanding the TV show library, as it currently lags behind the number of movies available.

2. Continue to offer a diverse range of genres, with a focus on International Movies and TV Shows, Dramas, and Comedies.

3. Explore opportunities to produce or acquire content in regions like the USA, India and the United Kingdom, as they have shown significant demand.

4. Offer a variety of movie durations, including those falling within the popular 90-100 minutes range.

5. Analyse and cater to the preferences of viewers by focusing on producing and promoting content in genres and ratings that have proven to be popular.

6. Highlight long-running TV shows with multiple seasons, like 'Grey's Anatomy,' 'Supernatural,' and 'NCIS,' to attract binge-watching audiences.