Netflix Movies Data Analysis Report

1. Executive Summary

The objective of this report is to analyze the Netflix movies dataset to understand content trends, audience preferences, and production distribution. The dataset comprises 9,827 movie entries. Prior to analysis, the data underwent rigorous cleaning and statistical transformation to ensure accuracy in evaluating key metrics such as Popularity and Genres.

2. Data Preprocessing & Cleaning Strategy

To derive accurate insights, the raw data was refined through the following steps:

- **Data Overview:** The initial dataset contained 9,827 rows and 9 columns. The data was found to be relatively tidy with no initial missing values or duplicates.

- **Date Standardization:** The 'Release_Date' column was converted into a datetime format. Subsequently, the specific 'Year' was extracted to facilitate the analysis of annual production trends.

- **Noise Reduction:** To maintain analytical focus, irrelevant columns such as 'Overview', 'Original_Language', and 'Poster_Url' were dropped as they did not contribute to statistical insights.

Advanced Categorization:

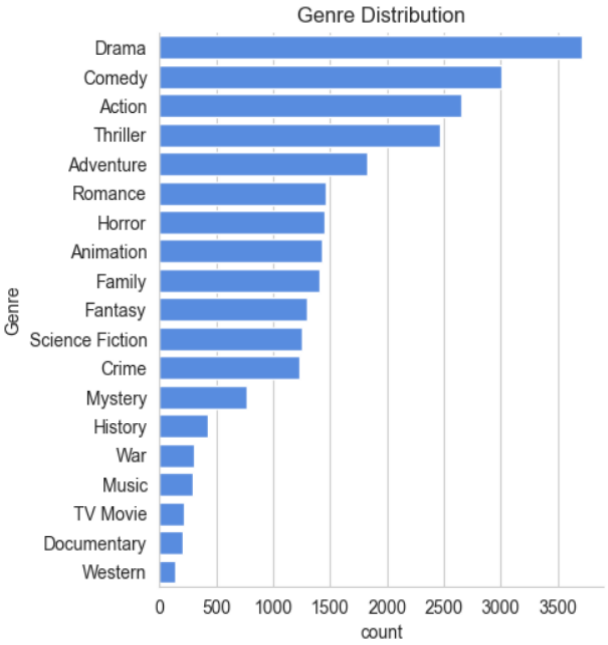
- **Vote Segmentation:** Instead of treating 'Vote_Average' purely as a raw number, it was segmented into four distinct categories: Not_Popular, Below_Avg, Average, and Popular. This segmentation was based on statistical quartiles (Min, 25%, 50%, 75%, Max).

- **Genre Expansion:** The 'Genre' column contained multiple genres per movie separated by commas. To accurately count genre frequency, the dataset was "exploded," creating a separate row for every genre associated with a movie.

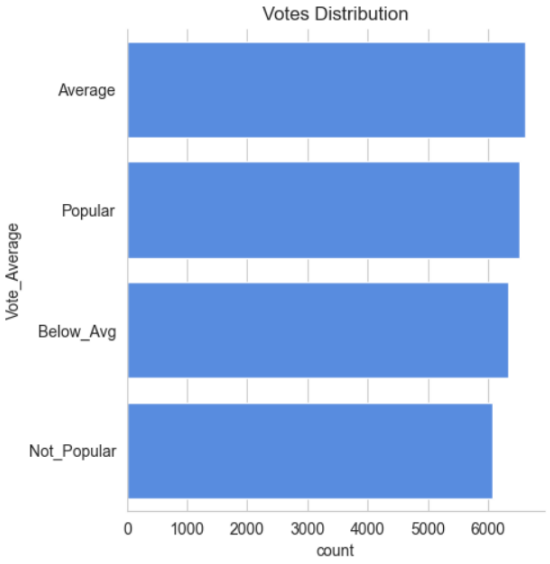
3. Detailed Analysis & Insights

3.1. Content Strategy: Most Frequent Genres

Question: What is the most frequent genre of movies released on Netflix?



Analysis: After splitting and exploding the genre data, the frequency analysis reveals that Drama is the most dominant genre on the platform. This indicates a strong strategic emphasis on storytelling and emotionally engaging content.



3.2. Audience Engagement: Vote Categories

Question: Which vote category dominates the Netflix movie dataset?

Analysis: The movies were analyzed based on their categorized vote averages. This distribution highlights which segment (Popular vs. Average) constitutes the majority of the content library.

3.3. Popularity Extremes: Highest vs. Lowest

This section identifies the outliers in the dataset—movies that performed exceptionally well or poorly.

🏆 **Highest Popularity:**

Movie: Spider-Man: No Way Home.

Genres: Action, Adventure, Science Fiction.

Insight: Audiences show a distinct preference for high-energy, visually rich, and franchise-driven content.

📉 **Lowest Popularity:**

Movies: "The United States vs. Billie Holiday" and "Threads".

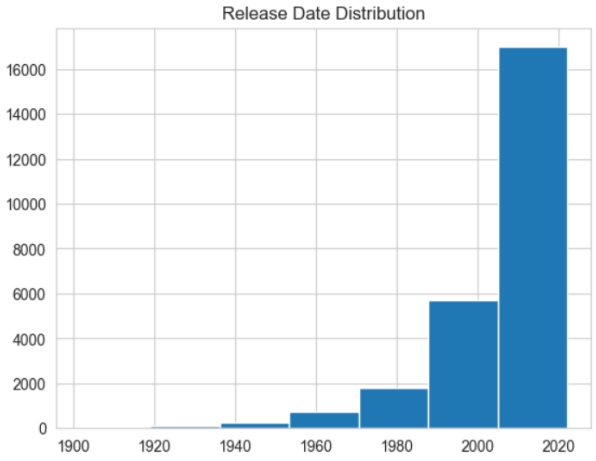
Genres: Music/Drama/History and War/Drama/Science Fiction respectively.

Insight: Movies with the lowest popularity tend to belong to niche genres or feature heavy/dark themes, which often have limited mass audience appeal.

3.4. Temporal Trends: Production Volume

Question: In which year were most movies filmed?

Analysis: The histogram of release years demonstrates a significant surge in content production in recent times.



Key Finding: The year 2020 recorded the highest number of movies filmed in this dataset, reflecting Netflix's aggressive content expansion strategy during that period.

4. Business Conclusion

- **Content Dominance:** Drama leads in volume, but Action/Sci-Fi leads in peak popularity.
- **Engagement Strategy:** To maximize engagement, the platform should balance its heavy investment in Drama with high-budget Action titles that drive massive popularity spikes.
- **Growth Trajectory:** The concentration of releases around 2018–2020 confirms a strategic pivot toward ramping up original content production.