NETFLIX MOVIE DATA ANALYSIS REPORT

VISHANKI

28.01.2025

Objective of the Data Analysis

The goal of this analysis is to extract meaningful insights from Netflix's movie dataset by examining trends in genres, release years, ratings, and country-wise content distribution. By leveraging data analytics techniques, we aim to understand Netflix's content strategy, identify key patterns, and provide recommendations for better decision-making in content acquisition and user engagement.

Dataset Overview

The dataset used in this analysis contains information about movies available on Netflix. The key attributes include:

- **Title:** The name of the movie.

- **Release Date:** The year the movie was released.

- **Genre:** The category or categories associated with the movie.

- **Duration:** The length of the movie in minutes.

- **Ratings:** The classification assigned based on audience suitability.

- **Country:** The country where the movie was produced.

The dataset provides a comprehensive overview of Netflix's content library and allows us to Analyze trends over time.

Data Cleaning & Preprocessing

Before performing any analysis, following steps were taken to clean and preprocess the data:

1. **Handling Missing Values:**

    - Identified and addressed missing values in the Genre, Country, and Ratings columns.

    - Filled missing values with 'Unknown' where applicable.

2. **Standardizing Date Formats:**

    - Converted Release Date into a standardized datetime format to ensure consistency.

3. **Splitting Multiple Genres:**

   o Some movies belong to multiple genres (e.g., "Action, Adventure"). These were split into separate rows to facilitate better analysis.

4. **Removing Duplicates:**

   o Checked and removed any duplicate movie entries to ensure data accuracy.
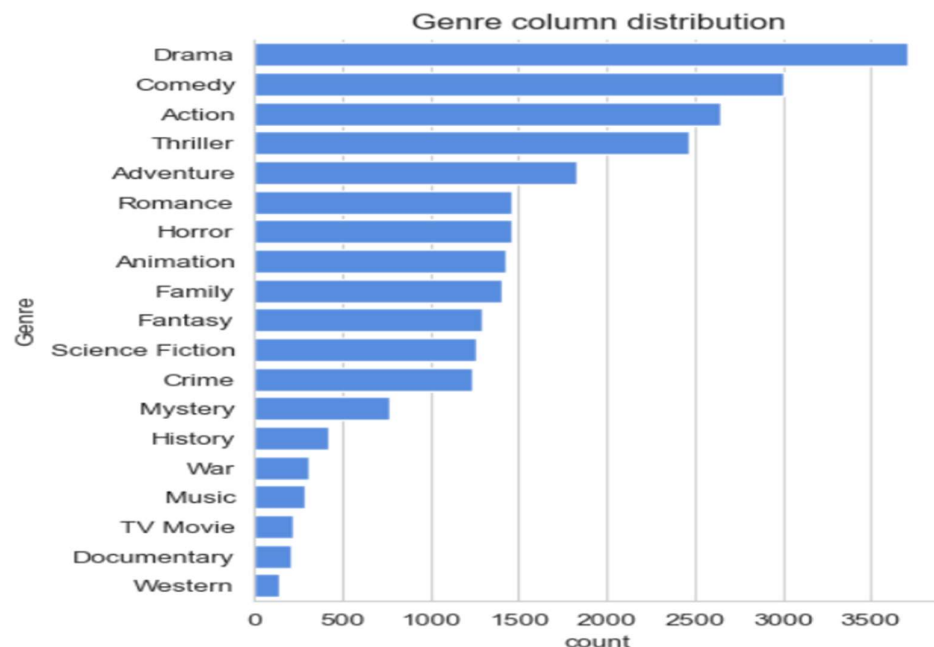
5. **Data Type Corrections:**

   o Converted numeric columns (like Duration) into appropriate integer formats for better analysis.

Exploratory Data Analysis (EDA) and Visualizations

Data visualization helps in understanding trends and patterns within the dataset. The following analyses were conducted:

1. Genre Distribution

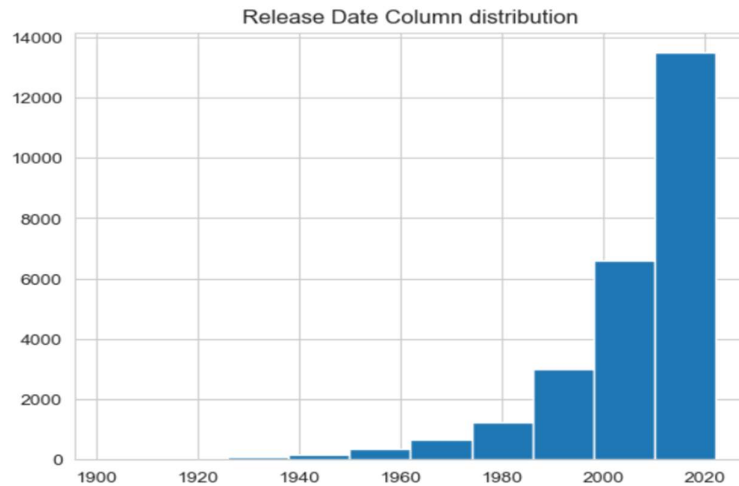- A bar chart was created to show the most common movie genres available on Netflix.

- **Insight:** Drama and Comedy are the most dominant genres.
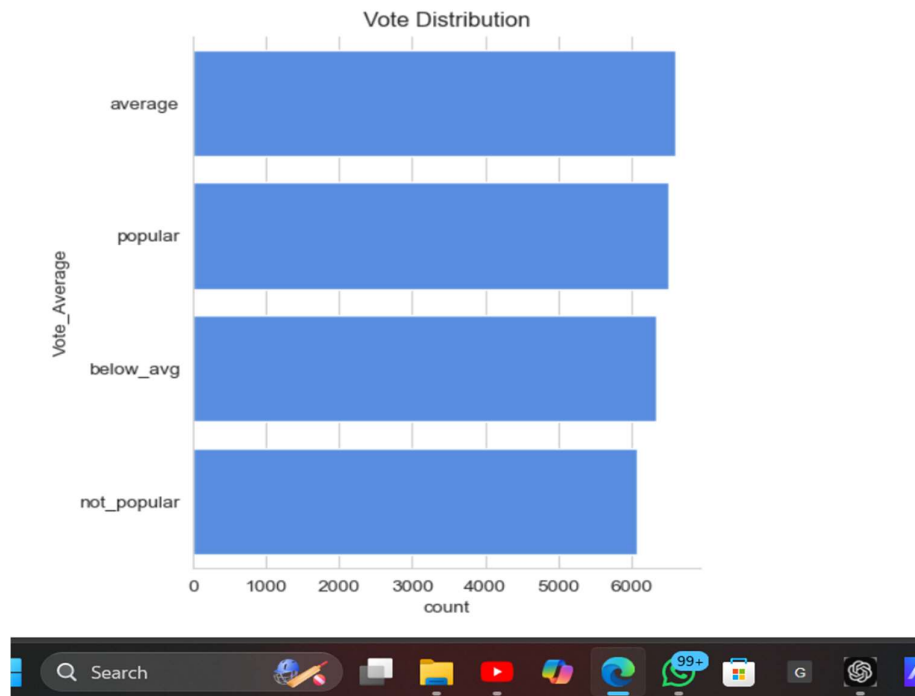
2. Yearly Movie Release Trends

- A line graph was used to display the number of movies released per year.

- **Insight:** There has been a significant rise in movie additions to Netflix's catalog in recent years.

```
In [36]:  ▶ df['Release_Date'].hist()
            plt.title('Release Date Column distribution')
            plt.show()
```

Release Date Column distribution

3. Ratings Analysis

- A histogram was plotted to analyze the distribution of movie ratings.

- **Insight:** TV-MA is the most common rating, indicating a focus on mature audience content.

Vote Distribution

4. Duration Analysis

- A box plot was used to examine the distribution of movie durations across different genres.

- **Insight:** Action and Drama movies tend to have longer durations compared to Comedy films.

Highest Popularity Movie and its Genre

```
In [34]: ▶ df[df['Popularity']==df['Popularity'].max()]
```

Out[34]:

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre
0	2021	Spider-Man: No Way Home	5083.954	8940	popular	Action
1	2021	Spider-Man: No Way Home	5083.954	8940	popular	Adventure
2	2021	Spider-Man: No Way Home	5083.954	8940	popular	Science Fiction

Least Popularity MoviE and its Genre

```
In [35]: ▶ df[df['Popularity']==df['Popularity'].min()]
```

Out[35]:

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre
25546	2021	The United States vs. Billie Holiday	13.354	152	average	Music
25547	2021	The United States vs. Billie Holiday	13.354	152	average	Drama
25548	2021	The United States vs. Billie Holiday	13.354	152	average	History
25549	1984	Threads	13.354	186	popular	War
25550	1984	Threads	13.354	186	popular	Drama
25551	1984	Threads	13.354	186	popular	Science Fiction

Key Insights

From the above analyses, we derived several important insights:

- **Content Growth:** Netflix's movie library has been expanding significantly over the years, with a rise in new additions.

- **Genre Preferences:** Drama and Comedy are the most dominant genres, aligning with audience demand.

- **Viewer Preferences:** TV-MA is the most frequent rating, suggesting a focus on mature content.

- **Movie Duration Trends:** Different genres have varying average durations, with Action movies typically being longer.

Final Insights & Conclusion

This analysis provides valuable insights into Netflix's movie catalog and its content distribution strategy. The findings indicate that Netflix focuses on Drama and Comedy while expanding its library significantly over the years. The dominance of TV-MA-rated content suggests a shift towards more mature content for audiences. Future analysis could focus on audience engagement, user ratings, and regional content preferences to enhance Netflix's recommendation algorithms and improve user satisfaction.