Netflix Movie Data Analysis Project

Introduction

Netflix, a leading streaming platform, leverages **data science** and **machine learning** to build strong recommendation systems and improve user experience. This project simulates a real-world data-driven analyst role, where the goal is to analyze a dataset of over 9,000 movies to provide actionable insights to aid business decision-making.

Objectives

- Determine the most **frequent genre** in Netflix's movie portfolio.
- Identify genres with the **highest audience votes**.
- Find the **movie with the highest popularity** and its genre(s).
- Find the **movie with the lowest popularity** and its genre(s).
- Discover which **year had the most films released**.
- Summarize findings and provide general visualization-based

Data Description

The dataset (mymoviedb.csv) consists of 9,827 movie records with the following columns:

- Release_Date: Movie release date (later processed as year)
- Title: Movie title
- Overview: Movie synopsis (not used in final analysis)
- Popularity: Numeric measure of popularity
- Vote_Count: Number of user votes
- Vote_Average: Average user rating
- Original_Language: Language (not used in final analysis)

- Genre: Comma-separated list of genres per movie

After cleaning and preprocessing, the dataset was expanded to over 25,000 rows, with each row representing a unique movie-genre combination, ensuring that each genre is counted separately for multi-genre movies.

Methodology

1. **Data Cleaning & Transformation:**
    - Removed Overview, Original_Language, and Poster_Url columns.
    - Converted Release_Date to extract the release year for annual analysis.
    - Handled missing/duplicated values (none found).
    - Split comma-separated genres into individual entries via dataframe expansion.
2. **Vote Categorization:**
    - Categorized Vote_Average into four bins: not_popular, below_avg, average, and popular using quartiles.
3. **Exploratory Data Analysis:**
    - Used Pandas, Matplotlib, and Seaborn for data aggregation and visualization.
    - Generated descriptive statistics and visual plots to answer each objective question.
4. **Genre Handling:**
    - Genres were exploded to analyze each genre distinctly.
    - Genre column was encoded as a categorical variable for efficient grouping and plotting.**Insights**

Most Frequent Genre

- The most frequent genre across all movies is **Drama**. It appears in over 14% of all genre entries, making it the dominant genre on Netflix.

Genres with Highest Votes

- The labeled popular vote category accounts for about 25.5% (~6520) of entries.

- **Drama** again is the genre most commonly associated with high popularity and user votes, representing more than 18.5% in the popular category

Movie with Highest Popularity

- **"Spider-Man: No Way Home"** scored the highest popularity rating in the dataset. Its associated genres are **Action, Adventure, and Science Fiction**.

Movie with Lowest Popularity

- Several records share the lowest popularity score (13.354). Titles include **"The United States vs. Billie Holiday"** (genres: Music, Drama, History) and **"Threads"** (genres: War, Drama, Science Fiction

Year with Most Movies Filmed

- The year **2020** had the highest number of movie releases in the dataset, which is reflected by a visible peak in the distribution of movie counts by year.

Visualizations Summary

- **Genre Distribution**: Bar chart illustrating Drama as the top genre, followed by others like Comedy and Action.
- **Vote Average Distribution**: Bar plot showing the breakdown of movies in categories from not_popular to popular.
- **Popularity Extremes**: Tabular/lookup visualization for movies with maximum and minimum popularity, including their genres.
- **Release Year Histogram**: Shows the sharp increase in movie counts, peaking at 2020, indicating a trend (may be partly due to increased direct-to-streaming releases during the pandemic).

Conclusion

- **Drama** is the most prevalent and most "popular" genre on Netflix, according to both frequency and vote-based metrics.
- The most popular movie is **"Spider-Man: No Way Home"** (Action/Adventure/Sci-Fi), while several movies share the lowest popularity scores, spanning genres from Music to War.
- The **year 2020** had the max number of movie releases, potentially reflecting shifts in entertainment distribution.

This project demonstrates a comprehensive, reproducible approach to Netflix movie data analysis. By employing robust data cleaning, categorization, and visualization, all key business questions were answered with statistical and visual rigor, yielding actionable insights for content planning and acquisition strategies.