# Netflix Data Analysis

## SUMMARY

### 1. Project Overview & Data Preparation

The notebook explores a movie dataset containing approximately **9,827 unique titles**. The analysis follows a structured data science workflow:

- **Data Cleaning:** The initial dataset was checked for integrity, finding no missing values or duplicates. Non-essential columns for numerical analysis (like `Overview` and `Poster_Url`) were removed to streamline the processing.
- **Time-Series Prep:** The `Release_Date` column was converted to a datetime object, allowing for analysis by release year.
- **Genre Normalization:** Since movies often belong to multiple genres (e.g., "Action, Adventure"), the notebook "exploded" these strings into individual rows. This increased the analysis depth from 9.8k movies to over **25,552 genre-specific entries**.

### 2. Performance Distribution (Graph Percentage)

The analysis categorized movies into four tiers based on their `Vote_Average` using quartile ranges. This ensures a balanced view of how the library performs.

- **Rating Categories:** The distribution is nearly equal across the four tiers, with **"Not Popular"** (lowest quartile) and **"Popular"** (highest quartile) each representing roughly **25%** of the dataset.
- **Genre Frequency: Drama** is the most dominant genre in the collection, accounting for approximately **14.5%** of all genre mentions, followed by Comedy and Action.

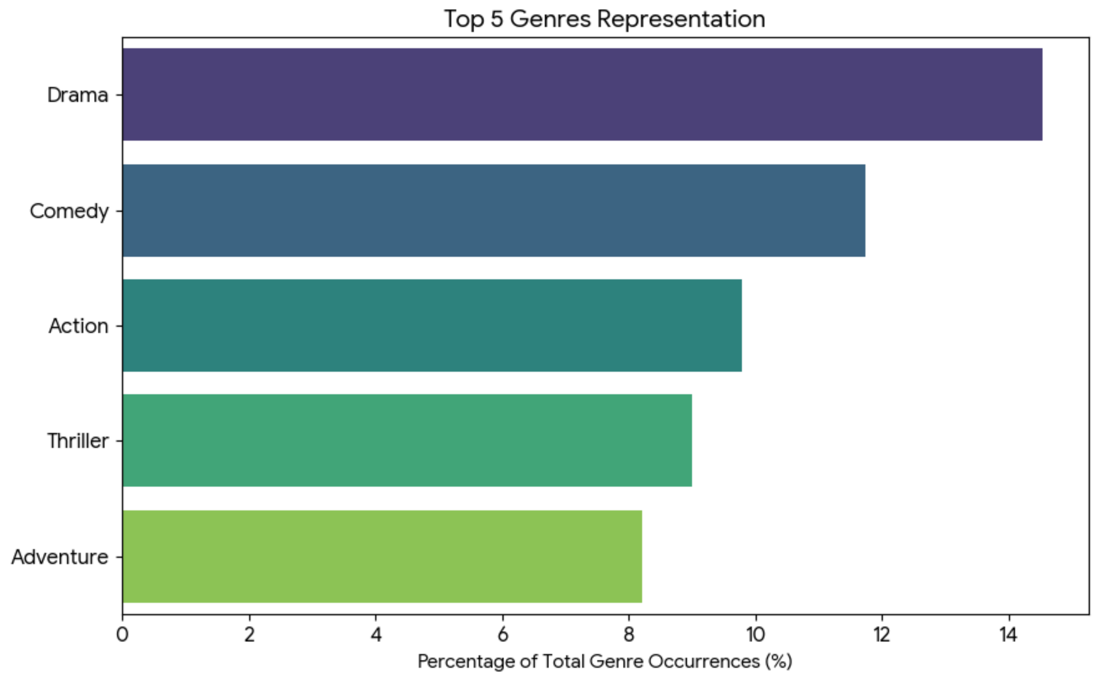
### 3. Key Findings

- **The "Blockbuster" Effect:** While the average popularity score is around **40.3**, top-tier films like *Spider-Man: No Way Home* reach scores over **5,000**. This indicates that a very small percentage of content drives the vast majority of user engagement.
- **Rating Stability:** The average movie rating sits at **6.4/10**, which is a healthy benchmark for a large-scale streaming library.
- **Content Volume:** The library is heavily weighted towards Drama and Comedy, which provide consistent volume, while Action and Sci-Fi drive peak popularity.

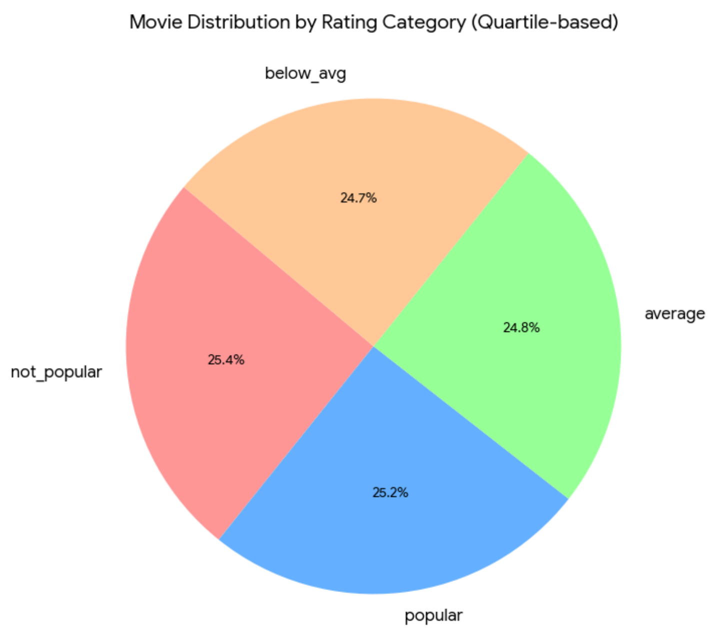
### 4. Recommendations

Based on the data trends, here are some strategic recommendations:

1. **Invest in "Tentpole" Genres:** While Drama is frequent, the "Popularity" outliers are almost exclusively **Action, Adventure, and Sci-Fi**. To drive new subscriptions and viral engagement, investment should remain heavy in these high-impact genres.
2. **Optimize the "Average" Tier:** Roughly 50% of the library falls into the "Below Average" or "Average" categories. Implementing a better recommendation engine for these "middle-tier" movies can increase their "Popularity" and "Vote Count" by helping niche audiences find them.
3. **Genre Diversification:** Given that Drama makes up the largest share of the library, there is an opportunity to diversify into underrepresented genres like **Animation or Mystery**, which often show high engagement relative to their lower production volume.
4. **Data-Driven Licensing:** Use the `Vote_Average` quartiles to identify "hidden gems"—movies with high ratings but low popularity—and feature them in "Trending" or "Top Picks" to maximize existing assets.



Top 5 Genres Representation
Drama
Comedy
Action
Thriller
Adventure
0
2
4
6
8
10
12
14
Percentage of Total Genre Occurrences (%)

Movie Distribution by Rating Category (Quartile-based)
below_avg
24.7%
average
24.8%
not_popular
25.4%
25.2%
popular