

Netflix Data Analysis Project

Netflix is known for its work in data science, AI, and ML, particularly for building strong recommendation models and algorithms that understand customer behavior and patterns. Suppose you are working in a data-driven job role, and you have a dataset of more than 9,000 movies. You need to solve the following questions to help the company make informed business decisions accordingly.

1. What is the most frequent genre of movies released on Netflix?
2. What genres has highest votes?
3. What movie got the highest popularity? what's its genre?
4. What movie got the lowest popularity? what's its genre?
5. Which year has the most filmed movies?

Project Overview

This project analyzes a dataset of ~9,000+ movies on Netflix to uncover trends in **genres, popularity, votes, and release years**. The goal is to support **data-driven decisions** using **data science and visualization techniques**

Dataset:

- Source: mymoviedb.csv
- Key columns used:
 - Title → Movie name
 - Genre → One or more genres per movie
 - Vote_Average → Average audience rating
 - Popularity → Popularity score
 - Release_Date → Release year

Data Cleaning Steps

- Converted Release_Date → Extracted **year only**.
- Dropped unnecessary columns (Overview, Original_Language, Poster_Url).
- Handled missing values by removing rows with NaN.
- Split Genre column into individual rows using `.explode()`.
- Converted Genre into a categorical variable

Business Questions & Insights:

1. What is the most frequent genre of movies released on Netflix?

- **Drama** and **Comedy** dominate as the most frequently released genres.
- These genres attract the widest audience, making them Netflix's core content categories.

2. Which genres have the highest votes?

- Genres like **Drama** and **Action** receive the highest number of votes.
 - Indicates audience engagement is strongest in story-driven and thrilling content.
-

3. What movie got the highest popularity? What's its genre?

- Dataset check shows one standout movie with **maximum popularity score**.
 - Its genre aligns with Netflix's global hit trends (often **Action/Adventure**).
-

4. What movie got the lowest popularity? What's its genre?

- At least one movie had **minimum popularity**, usually from a **niche genre**.
 - This shows that not all genres perform equally well, and audience tastes vary widely.
-

5. Which year has the most filmed movies?

- Histogram of Release_Date shows a **peak around recent years (2017–2019)**.
 - Suggests Netflix's movie production grew rapidly before the pandemic period.
-

Visualizations:

- **Genre Distribution** → Bar chart showing Drama & Comedy on top.
 - **Votes Distribution** → Ratings mostly cluster in mid-range values.
 - **Popularity Extremes** → Identified max & min movies.
 - **Release Year Trend** → Histogram confirming surge in recent years.
-

Conclusion:

- **Drama & Comedy** are Netflix's most frequent genres.
- **Drama & Action** genres receive the highest votes, reflecting popularity.
- **One movie leads in popularity** while some niche films lag behind.
- **2017–2020 had the most movie releases**, marking Netflix's aggressive expansion.

This analysis helps Netflix (or similar platforms) understand **content trends** and guide **future production strategies**.