



# Netflix Movie Data Analysis

**By:** Sushmita Sahu | **Role:** Data Analyst

**Tools:** Python | Pandas | Seaborn | Matplotlib



A comprehensive data analysis project exploring Netflix's movie catalog to uncover patterns in genre popularity, audience preferences, and content release trends.

# The Netflix Story: From DVDs to Streaming Giant

Netflix was founded on **August 29, 1997**, in Scotts Valley, California, by Reed Hastings and Marc Randolph. What began as a DVD-by-mail service transformed into the world's leading streaming entertainment platform.

Hastings, a computer scientist and mathematician, was inspired to create Netflix after receiving a hefty late fee from a traditional video rental store. Randolph, a marketing executive, shaped the company's early user interface and distinctive branding that we recognize today.





# Project Overview



## Primary Objective

Analyze Netflix movie data to uncover actionable insights about genre preferences, popularity metrics, and content release patterns over time.



## Key Analysis Goals

- Identify the most frequent and popular genres
- Discover top-rated and least popular movies
- Track release trends and peak production years



## Business Value

Provide data-driven insights to understand viewer preferences, guide content strategy decisions, and inform future production planning.

# Dataset Information & Structure

## Dataset Specifications

- **Source:** iscale Netflix Movie Dataset
- **Initial Size:** 9,827 rows × 9 columns
- **Cleaned Size:** 6 essential columns retained
- **Data Quality:** No missing or duplicate values detected

## Key Columns Analyzed

- Title
- Genre
- Popularity Score
- Vote Count
- Vote Average
- Release Year

## Data Observations

- ❑ **Outliers Detected:** The Popularity column contained significant outliers that required careful handling during analysis to prevent skewed results.

- ❑ **Genre Complexity:** Movies contained multiple genres separated by commas, requiring specialized cleaning and transformation techniques.



# Data Cleaning & Transformation

01

---

## Date Conversion

Converted Release\_Date field to Year format (integer type) for streamlined temporal analysis and trend visualization.

02

---

## Column Reduction

Dropped unnecessary columns: Overview, Original\_Language, and Poster\_URL to focus analysis on core metrics.

03

---

## Vote Categorization

Categorized Vote\_Average into four distinct segments: *not\_popular*, *below\_avg*, *average*, and *popular* for clearer insights.

04

---

## Genre Separation

Split multiple genres into individual rows using the explode() function, enabling precise genre-level analysis.

**Final Transformed Dataset:** 25,552 rows × 6 columns — ready for comprehensive exploratory analysis.



## Key Findings: Genre Dominance & Popularity

14%

### Drama Genre Frequency

Drama is the most frequent genre, appearing in 14% of all movies across 19 total genres in the dataset.

18.5%

### Drama Popularity Share

Drama dominates with 18.5% of total popularity scores, confirming its position as the audience favorite.

25.5%

### Popular Vote Category

Approximately 6,520 movies (25.5%) fall into the "popular" vote category, indicating strong audience engagement.

### 🕸 Most Popular Movie

**Title:** *Spider-Man: No Way Home*

**Genres:** Adventure, Science Fiction

**Popularity Score:** 5083.954

### 📉 Least Popular Movies

**Titles:** *The United States vs. Billie Holiday* and *Threads*

**Genres:** Music, Drama, War, Sci-Fi, History

**Popularity Score:** 13.354

# Release Trends & Production Peak

## 🎬 Year with Most Movie Releases

The year **2020** recorded the highest number of movie releases in the entire dataset, marking a significant milestone in Netflix's content expansion strategy.

This surge coincides with the global shift toward streaming entertainment and increased production investment during the pandemic era.



## Tools & Technologies

### Programming

Python (Pandas, NumPy)

### Visualization

Matplotlib, Seaborn

### Environment

Jupyter Notebook



## Conclusion & Strategic Insights



### Drama Dominance

Drama emerges as both the most frequent and most audience-favored genre, representing a reliable content strategy foundation.



### Blockbuster Winner

*Spider-Man: No Way Home* achieved unprecedented popularity scores, showcasing the power of franchise content.



### 2020 Production Peak

Film production reached its zenith in 2020, reflecting evolving content consumption patterns and streaming growth.

**Business Impact:** This analysis provides valuable insights into [viewer preferences](#) and [content trends](#), empowering data-driven decisions for content acquisition, production planning, and audience targeting strategies.

Understanding these patterns helps Netflix and content creators optimize their portfolios, allocate resources effectively, and deliver content that resonates with diverse global audiences.