1. Title Page:

Project Title: Netflix Data Analysis
Name: Mihir Aryan Mishra
Tools Used: Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook
Date: 27th Oct, 2025

2. Introduction:

This project analyzes Netflix's content catalog to explore viewing trends, genre popularity, and content distribution by country and year.

The aim is to extract meaningful insights that reflect how Netflix's global content strategy evolved over time.

3. Objectives:

List specific goals, such as:

- Determine the total number of titles available on Netflix.

- Compare the count of **Movies vs TV Shows**.

- Identify **most popular genres** and **countries of production**.

- Analyze **content rating distribution** (e.g., TV-MA, PG, etc.).

- Examine how content additions evolved **year-over-year**.

4. Dataset Description:

Include dataset details like:

- **Source:** Netflix dataset from Kaggle.

- **Size: ~8,800 records, 12 columns.**

- **Columns Explained:**

  - **title, type, country, release_year, rating, duration, etc.**

- **Mention any data cleaning done:**

  - Removed duplicates and nulls.

  - Filled missing country or director fields if possible.

  - Standardized column names.

5. Tools & Libraries Used:

List everything you used:

- Python (Pandas, NumPy)

- Matplotlib, Seaborn (for data visualization)

- Jupyter Notebook (for execution)

- Excel / CSV (data preprocessing)

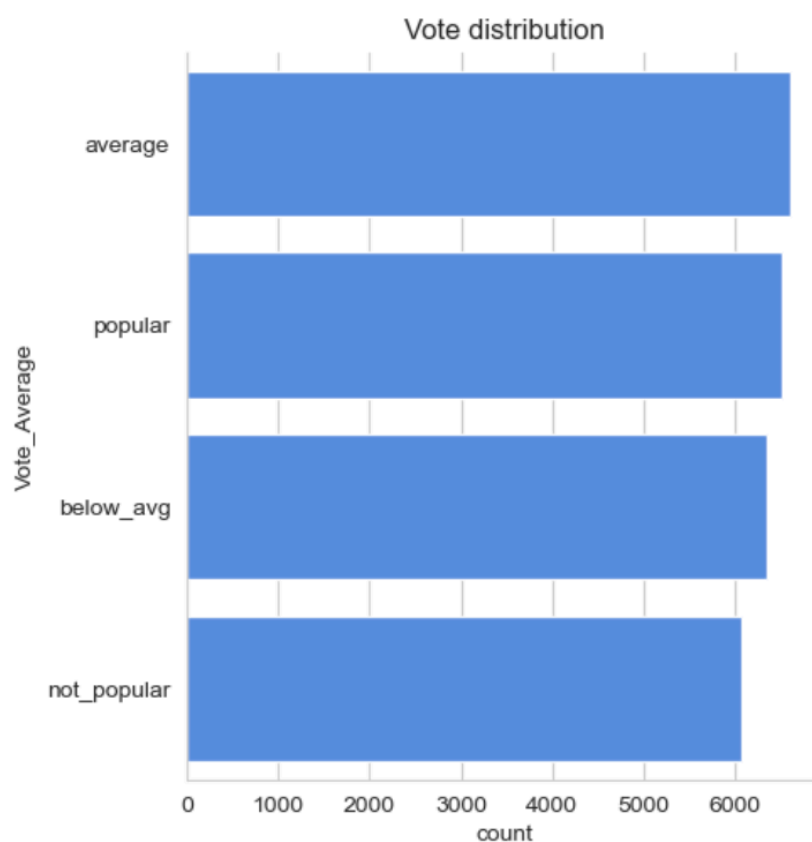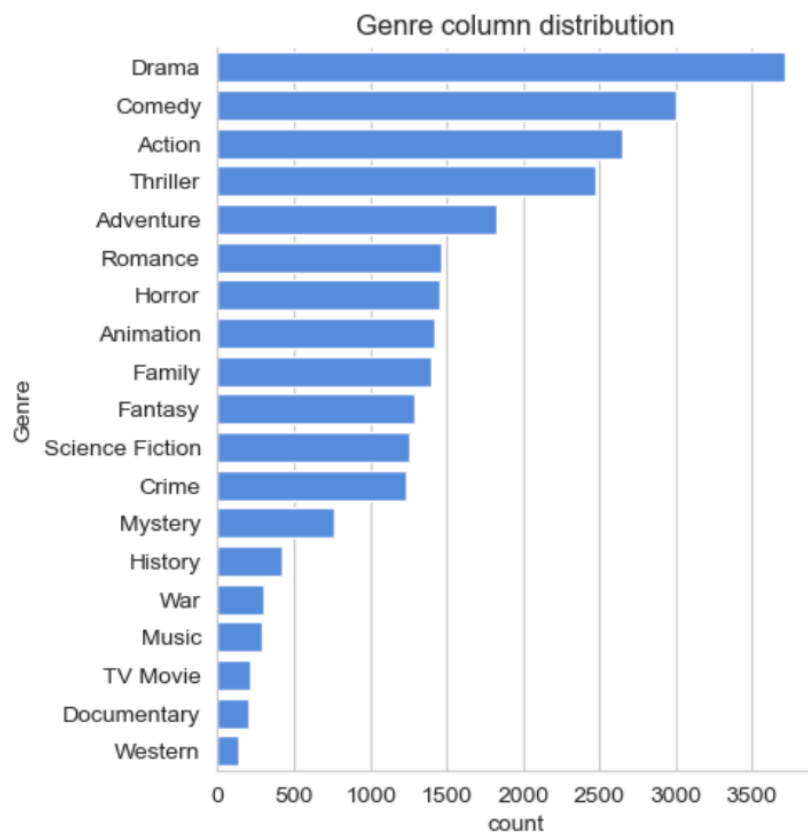## 6. Data Cleaning & Preprocessing

Describe what you did before analysis:

- Checked for missing values and duplicates.

- Converted date_added to datetime format.

- Extracted year and month features if needed.

- Filtered out irrelevant or invalid entries.

- Example: "Out of 8,800 entries, 130 were duplicates and removed."

7. Exploratory Data Analysis (EDA)

Include analysis like:

- **Type Distribution:** Movies vs TV Shows (Bar Chart)

- **Country-wise Content:** Top 10 countries with most titles

- **Genre Trends:** Top 10 genres or "listed_in" keywords

- **Release Year Trends:** Titles added by year

- **Ratings:** Most frequent rating categories (TV-MA, TV-14, etc.)

- **Directors/Actors:** Most frequent contributors

Genre column distribution

Vote distribution

8. Key Insights & Findings

Movies make up **~70%** of Netflix content.
USA and **India** are the top producers.
Drama and **Comedy** dominate all genres.
The majority of content targets mature audiences (TV-MA).
Peak content addition years: **2017–2020**.

9. Results Summary Table

Metric	Insights
Total Titles	8,800+
Dominant Type	Movies (70%)
Top Genres	Drama, Comedy
Top Countries	USA, India
Peak Years	2017–2020
Common Ratings	TV-MA, TV-14

10. Future Work / Recommendations

Build an interactive Power BI or Tableau dashboard.
Perform sentiment analysis on movie descriptions.
Predict user preferences using machine learning.
Compare Netflix data with competitors (Prime, Disney+) for deeper insights.

11. Conclusion

The Netflix Data Analysis highlights the platform's rapid expansion, diverse content strategy, and preference for mature-rated titles.

The analysis provides a foundation for understanding audience demand and could guide recommendations for regional or genre-based investments.