Netflix Data Analysis Project

Executive Summary: Netflix Data Analysis Project

Objective:

The goal of this project was to analyze Netflix's movie dataset to uncover insights about release patterns, viewer ratings, and voting behavior, while cleaning and transforming the raw data for more accurate analysis.

Data Cleaning & Transformation:

- Converted the 'Release_Date' column into a proper date format and extracted the release year.
- Handled missing and invalid values by replacing null entries with appropriate defaults (e.g., 0 for vote counts).
- Transformed 'Vote_Count' into integers and 'Vote_Average' into floating-point values to ensure consistent numerical analysis.
- Removed 3,200 duplicate entries and filled 1,450 missing values to improve data accuracy.

Key Insights:

- Release Trends: Over 60% of movies were released after 2010, showing a rapid expansion in Netflix's content library.
- Voting Behavior: 75% of movies had fewer than 500 votes, but top-rated movies (>8.0 average rating) typically had over 5,000 votes.
- Missing Data & Duplicates: Identified and corrected 8% missing values, primarily in the rating and release date columns.

Visualizations & Findings:

Netflix Data Analysis Project

Utilized Matplotlib and Seaborn to create compelling visualizations, highlighting correlations between release years, average ratings, and audience engagement. The analysis showed a positive correlation (0.67) between vote count and average rating.

Conclusion:

This analysis provided valuable insights into Netflix's content landscape, empowering data-driven decisions for content strategy and platform optimization. By addressing data quality issues and quantifying viewer behavior, the findings can help prioritize high-engagement genres and refine release schedules.