Netflix Movies and TV shows EDA

Project Overview:
This project analyzed a Netflix dataset to uncover key insights about its movies and TV shows. The dataset contained over 8000 titles, with features such as title, director, cast, country, date_added, release_year, rating, genre, description, and duration.

The analysis focused on data wrangling, cleaning, visualization, and generating actionable insights to understand Netflix's content trends.

Key Areas of Analysis:

Data Wrangling and Cleaning:

1. Handling Missing Values:
    - Columns like director, cast, and country had missing values. Missing date_added entries were replaced with the earliest date in the dataset.
    - Normalization of categorical data, such as country names (U.S., USA → United States).
2. Standardization:
    - Text data, such as titles, was converted to title case (e.g., breaking bad → Breaking Bad).
    - Ratings with low frequencies (less than 5 occurrences) were grouped into an "Other" category.
3. Feature Engineering:
    - Extracted numeric values from the duration column for movies and stored them as integers.
    - For TV shows, the number of seasons was extracted and stored in a new column.

Insights from the Data:

1. Content Distribution:
    - Movies dominate the Netflix library, significantly outnumbering TV shows.
    - Visualized with bar charts and a donut chart, highlighting the proportion of movies (~70%) to TV shows (~30%).
2. Content Trends Over Time:
    - A dramatic rise in content production was observed after 2000, indicating Netflix's expansion strategy.
    - Before 2000, the number of releases remained consistently low.
3. Genre and Country Analysis:
    - Drama, Comedy, and Documentary are the most prominent genres.
    - The United States leads content production with over 2,800 titles, far ahead of other countries such as India and the UK.
4. Ratings Distribution:
    - Analyzed the distribution of ratings (e.g., PG, TV-MA, etc.) to identify audience preferences.
    - Grouped infrequent ratings into "Other" for better clarity in visualizations.

Data Visualizations:

- Bar and Pie Charts: Showed the distribution of content by type and proportions of movies vs. TV shows.
- Line Graphs: Illustrated the release trends over the years, revealing exponential growth post-2000.
- Horizontal Bar Charts: Depicted the top 10 countries by content production, highlighting U.S. dominance.

Key Findings:

1. Netflix's Growth:
    - A clear strategy of exponential content expansion, particularly after 2000, with a focus on original productions.

2. Regional Trends:
    - The U.S. leads in content production, but other countries are steadily contributing, especially in regional genres.
3. Audience Preferences:
    - Drama and Comedy continue to be crowd favorites, dominating Netflix's catalog.
    - Ratings indicate a diverse audience, but some gaps in content targeting younger demographics are evident.

Conclusion:

This analysis demonstrates how Netflix's data reflects its strategic growth, global reach, and efforts to cater to diverse audience preferences. From handling missing data to visualizing trends, this project provided hands-on experience in data analysis and helped uncover meaningful insights about one of the largest streaming platforms.