Strategic Decision Making with PowerBI

Prof Arpit Yadav

"Netflix Content Analysis Dashboard: Unlocking Trends with Power BI"

Submittied By: Kanchan Bhargava

2023JULB01178

Table of Content

1. Problem Statement

The purpose of this analysis is to explore **content trends on Netflix**, including insights into the types of content available (movies vs. TV shows), the genres offered, their ratings, and regional availability. The goal is to help decision-makers or analysts understand content production and viewer preferences to guide Netflix's strategic decisions. This includes:

- Identifying the most popular genres.
- Understanding trends in content production over time.
- Analyzing the distribution of content across countries.

---

2. Data Requirement

To address the problem statement, specific data fields from the dataset are necessary. These include:

- **Title**: The name of the content.
- **Type**: Whether it's a movie or TV show.
- **Release Year**: When the content was made available.
- **Rating**: Content classification based on age suitability (e.g., TV-MA, PG-13).
- **Duration**: The runtime of movies or the number of seasons for TV shows.
- **Country**: The primary country of production.
- **Genre**: The content category (e.g., Drama, Comedy).

This data will allow us to answer questions like which genres are most common, what types of content are being produced more frequently, and which countries are key contributors to Netflix's library.

---

3. Data Collection

The data was sourced from the Netflix dataset provided in .csv format. It contains several columns, including:

- **show_id**: A unique identifier for each show.
- **title**, **type**, **director**, **cast**, **country**, **date_added**, **release_year**, **rating**, **duration**, **genre**, and **description**.

- This dataset was likely created by scraping Netflix's publicly available data or using an API.

---

4. Data Validation

To ensure the data is accurate, we performed the following checks:

- **Completeness Check**: Verified that all required fields (e.g., title, type, rating) have values.
- **Consistency Check**: Checked for uniform data formats (e.g., dates in YYYY-MM-DD).
- **Duplicate Removal**: Identified and removed duplicate records using the **show_id** field.
- **Range Validation**: Verified that the release years are between 1925 and 2021, as shown in the dashboard.

---

5. Data Cleaning

This step involved:

- **Handling Missing Values**:
    - For columns like director or cast, missing values were replaced with "Unknown" or left blank if not critical.
    - Missing country data was categorized as "Worldwide."
- **Outlier Treatment**:
    - Extreme values in **duration** were reviewed (e.g., unusually long movie durations) and corrected or excluded if invalid.
- **Formatting**:
    - Ensured numerical fields (e.g., release_year) and categorical fields (e.g., rating) were properly formatted.

---

6. Tools

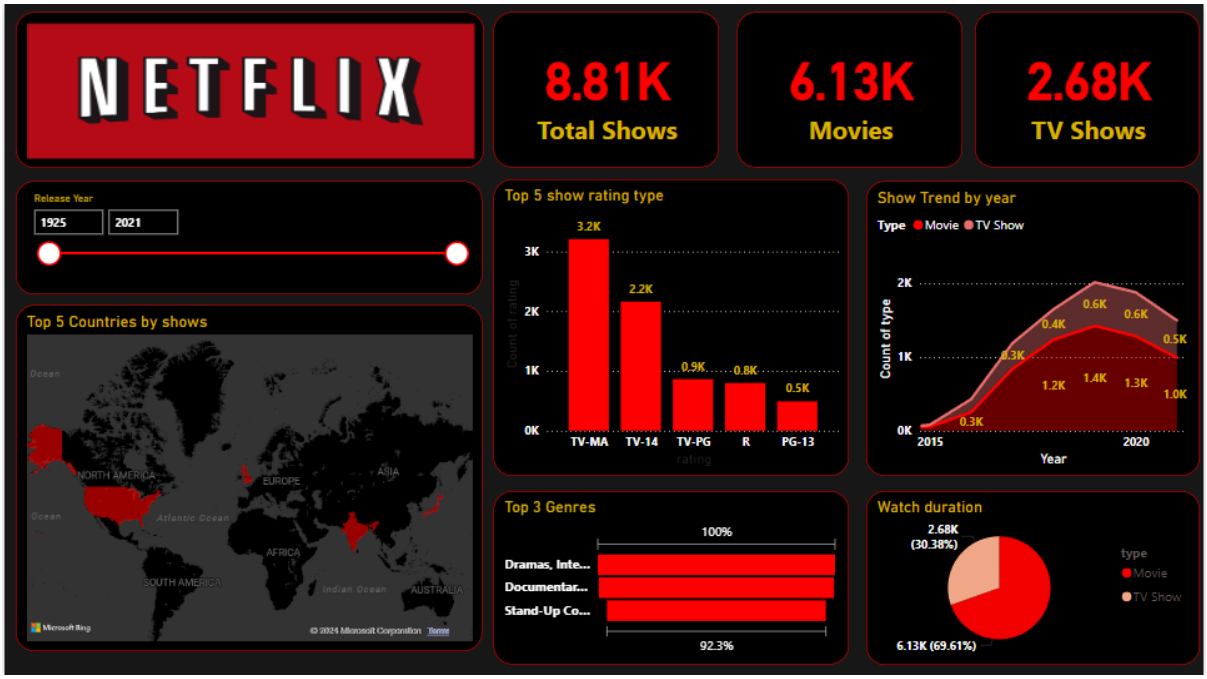
The analysis leveraged the following tools:

- **Power BI**: Used for creating interactive visualizations and the dashboard.

- **Excel**: Employed for initial data cleaning, exploration, and validation.
- **Tableau (optional)**: Another visualization tool that could replicate or enhance the dashboard.

---

7. Dashboard



The dashboard provides a visual summary of the data:

- **Definition**: A dashboard is a collection of visual elements that provide insights into data, enabling users to quickly interpret trends and metrics.
- **Importance**: Dashboards allow for real-time decision-making and concise representation of complex datasets.

Key Graphs in the Dashboard:

1. **Univariate Analysis**:
    - **Bar Chart for Ratings**: Shows the count of movies and TV shows for each rating (e.g., TV-MA, PG-13). This helps identify the dominant ratings on Netflix.
    - **Pie Chart for Watch Duration**: Highlights the proportion of movies vs. TV shows.
2. **Bivariate Analysis**:

- **Line Chart for Trends Over Time**: Compares the growth of movies and TV shows released between 2015-2021.
- **Geographical Map**: Displays the top 5 countries producing Netflix content, providing insights into regional contributions.

3. **Multivariate Analysis**:
    - Combined views of genres, ratings, and type to analyze patterns (e.g., which genres are popular in different countries or ratings).

---

8. Storytelling

What is Analytical Storytelling?

It involves using data visualizations and narratives to communicate insights clearly. Instead of just presenting raw data, storytelling combines visuals, context, and key findings to drive understanding and engagement.

Importance:

- Helps decision-makers understand complex data easily.
- Provides actionable insights by highlighting key trends and anomalies.
- Enhances retention of information through a logical flow of data.

Story from the Dashboard:

- Netflix's **content library has grown significantly** in recent years, especially from 2015 onward, with a higher focus on movies compared to TV shows.
- **Drama and International genres dominate**, making up over 92% of the library.
- Countries like **the USA, India, and the UK** are top contributors to Netflix's library, reflecting global content diversity.
- Ratings like **TV-MA and TV-14** are the most common, suggesting a focus on mature audiences.

---

9. Conclusion

The analysis shows that Netflix has rapidly expanded its content, focusing on movies and drama genres, with contributions from multiple countries. The dashboard provides a clear understanding of Netflix's trends, helping the company focus on areas like

increasing content for specific genres or improving representation in underrepresented regions.