Project Report: Netflix Content Analysis using Power BI

A Power BI Capstone Project Report

Submitted by:
Syam Kumar Appikatla

Mentor :
mrs.Siddhika shah

Project for:
Sure Trust Internship

Abstract

This report details a data analysis project focused on the Netflix content library. The primary objective was to use Microsoft Power BI to develop an interactive dashboard that provides insights into the composition of movies and TV shows available on the platform. By processing and visualizing a comprehensive dataset of Netflix titles, the project uncovers key trends and patterns related to content types, genres, viewer ratings, and movie durations. The analysis reveals that the Netflix library is predominantly composed of movies, with a strategic focus on content for mature audiences (TV-MA and TV-14 ratings). The most prominent genres include Dramas, International Movies, and Comedies. Furthermore, the data shows a clear preference for feature-length films (90+ minutes). The resulting dashboard serves as a powerful tool for understanding Netflix's content strategy and provides a foundation for data-driven decision-making in content acquisition and production.

1. Introduction

1.1 Background

In the highly competitive streaming market, content is the primary driver of subscriber acquisition and retention. Netflix, as a global leader, possesses a vast and diverse library of movies and TV shows. Understanding the composition of this library is crucial for making strategic decisions, identifying content gaps, and catering to audience preferences. Data analysis provides the tools to transform raw content data into actionable insights, helping stakeholders understand what is available on the platform and how it is categorized.

1.2 Problem Statement

The sheer volume of content on Netflix makes it challenging to get a clear, high-level overview of its library without proper analytical tools. There is a need to systematically

analyze and visualize key metrics—such as the distribution of movies versus TV shows, the most common genres, and the prevalence of different content ratings—to understand the platform's strategic focus.

1.3 Objectives

The main objectives of this project were:

- To design and build a dynamic, interactive dashboard using Microsoft Power BI.
- To analyze and visualize the distribution of movies and TV shows in the Netflix library.
- To identify the most prominent genres and content ratings on the platform.
- To examine the distribution of movie runtimes.
- To extract and present key insights that summarize Netflix's content strategy.

1.4 Scope

The scope of this project is centered on the analysis of a static Netflix dataset. It involves the entire process of data cleaning, transformation, and visualization within Power BI to create a multi-page dashboard. The analysis focuses on four primary areas: content type, content rating, genre, and movie duration.

2. Methodology and Tools

2.1 Primary Tool: Microsoft Power BI

Microsoft Power BI was the exclusive tool used for this project. It is a powerful business intelligence platform that enables end-to-end data analysis, from connecting to data sources to creating compelling, interactive visualizations. Its capabilities were leveraged for data transformation (using Power Query), data modeling, and the creation of the final dashboards.

2.2 Data Source

The project utilized a dataset containing details of Netflix titles. The key data fields included:

- `show_id`: Unique identifier for each title.
- `type`: Category (Movie or TV Show).
- `title`: Name of the movie or show.
- `country`: Country of production.
- `release_year`: The year the content was released.
- `rating`: The content rating (e.g., TV-MA, PG-13).
- `duration`: The runtime of the movie or number of seasons for a TV show.

- listed_in: The assigned genres.

2.3 Project Workflow

1. **Data Loading and Cleaning:** The dataset was imported into Power BI. The Power Query Editor was used to handle inconsistencies and missing values (e.g., nulls in the country field).
2. **Data Transformation:** New columns and measures were created to facilitate analysis. This included creating "Duration Buckets" (e.g., '0-30 min', '30-60 min', '60-90 min', '90+ min') to categorize movies based on their runtime.
3. **Data Visualization:** A variety of visuals were created to represent the findings, including:
    - **Key Performance Indicator (KPI) Cards:** To display top-line numbers like Total Titles, Total Movies, and Total TV Shows.
    - **Donut and Bar Charts:** To show the distribution and comparison of categorical data like content ratings and genres.
    - **Treemap:** To visualize the hierarchy and weight of different genres.
4. **Dashboard Design:** The visuals were organized into three distinct, user-friendly dashboard pages: an overall content overview, a detailed analysis page, and a final summary of key insights.

3. Data Analysis and Findings

The analysis of the Netflix dataset yielded several key findings, presented across the interactive dashboards.

3.1 Overall Content Library Composition

The dashboard reveals that the Netflix library is extensive, with a clear focus on one type of content.

- **Total Titles:** 8,792
- **Total Movies:** 6,126 (~70%)
- **Total TV Shows:** 2,664 (~30%)

This shows a significant strategic inclination towards movies, which outnumber TV shows by more than two to one.

3.2 Content Ratings Distribution

The analysis of content ratings highlights a clear target audience.

- **TV-MA (Mature Audience):** The most common rating, with **3,210 titles (36.5%)**.
- **TV-14 (Parents Strongly Cautioned):** The second most common, with **2,160 titles (24.5%)**.

- **TV-PG (Parental Guidance Suggested):** Accounts for **860 titles (9.8%)**.
- **R (Restricted):** Accounts for **800 titles (9.0%)**.

Collectively, content aimed at mature audiences (TV-MA, TV-14, R) makes up over 70% of the library, indicating a focus on adult and older teenage viewers.

3.3 Top Genres

The analysis of genres, based on the "listed_in" category, shows a focus on specific types of content.

- **Top Genre Combinations:**
  1. Dramas, International Movies: **362 titles**
  2. Documentaries: **359 titles**
  3. Stand-Up Comedy: **334 titles**
  4. Comedies, Dramas, International Movies: **274 titles**
  5. Dramas, Independent Movies, International Movies: **252 titles**

Dramas, Comedies, and Documentaries are the dominant high-level categories. The frequent appearance of "International Movies" in top combinations also underscores Netflix's global content strategy.

3.4 Movie Duration Analysis

The analysis of movie runtimes reveals a preference for traditional, feature-length films.

- **90+ min:** This is the largest category, with **4,137 movies**.
- **60-90 min:** The second-largest category.
- **30-60 min & 0-30 min:** These shorter formats are significantly less common.

This suggests that Netflix's movie acquisition and production strategy prioritizes the standard cinematic experience over short films or mid-length features.

4. Conclusion and Key Insights

This project successfully demonstrated the use of Power BI to analyze a large content library and extract meaningful, strategic insights. The dashboard provides a clear, multi-faceted view of the Netflix catalog.

Key Insights:

- **A Movie-Dominant Platform:** Netflix's content strategy is heavily weighted towards movies, which comprise roughly 70% of its entire library.
- **Targeting Mature Viewers:** With TV-MA and TV-14 being the top two ratings, Netflix's primary target audience is adults and older teens.

- **Drama is the Core Genre:** Dramas, often blended with international and comedy elements, form the backbone of the Netflix catalog, complemented by a strong offering of documentaries and stand-up comedy.
- **Focus on Feature Films:** The platform overwhelmingly favors standard feature-length films (90+ minutes), investing less in shorter-form movie content.

5. Future Scope

This project can be expanded in several ways to provide even deeper insights:

- **Time-Series Analysis:** Analyze trends over time by incorporating release dates to see how content strategy has evolved year over year.
- **Performance Metrics:** Integrate viewership data or audience ratings (e.g., from IMDb or Rotten Tomatoes) to measure the popularity and success of different types of content.
- **Competitive Analysis:** Expand the dataset to include content from other streaming platforms (e.g., Disney+, HBO Max) to perform a comparative analysis of content libraries.
- **Predictive Modeling:** Utilize the existing data to build a model that can predict which attributes (genre, cast, director, etc.) are most likely to correlate with a title's success.