**Project - Phase I: Planning – Team 23**

**Kartheek Panyam & Gokuleswar Reddy Mooli**

**Arun Pal Bommaji & Maneesh Reddy Devara**

MS IN INFORMATION TECHNOLOGY, ARIZONA STATE UNIVERSITY

**IFT 533: Data Visualization & Reporting for IT**

**PROF. ASMAA ELBADRAWY**

APRIL 20, 2025

**Planning:**

**1.** We analyzed a range of datasets from reputable sources like Kaggle. Our analysis covered three major datasets. First, the **Netflix Movies and TV Shows** dataset, which has 8,807 records and 12 columns, including attributes like title, genre, country, release year, and rating. Second, the **IMDb Top 1000 Movies** dataset, which has 1,000 records and 16 columns, including genre, runtime, rating, and gross income. Third, the **Spotify Music Data** data set includes over 160,000 rows and 23 columns, attributes of which are tempo, danceability, energy, and genre. We went through all the datasets for the kinds of attributes (categorical, interval, ratio, ordinal), the analysis types that other data scientists performed (e.g., genre trends, audio clustering, and rating vs. revenue), and what could be visualized from bar plots and line charts all the way through to heatmaps and network graphs. This enabled us to gauge the richness and usability of each dataset to produce an engaging and insight-led dashboard.

**2.** I have selected the "Netflix Movies and TV Shows" dataset for this project. It contains over 8,807 records of Netflix content, including both movies and TV shows. The dataset has 12 columns that describe each title with details such as title, type, director, cast, country, date added to Netflix, release year, rating (e.g., PG, TV-MA), duration, genre (listed\_in), and a short description.

This dataset is useful for building an entertainment analytics dashboard that can explore:

* Genre trends
* Year-wise content growth
* Rating distribution
* Movie vs. TV show proportions
* Regional content production (based on country)

It provides a clear picture of the kind of content Netflix offers and can help analyze viewing trends, target audiences, and content strategy.

**3**.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Explanation |
| show\_id | Categorical | Unique identifier, treated as labels |
| type | Categorical | Two categories: Movie or TV Show |
| title | Categorical | Names of titles; no order |
| director | Categorical | Names of directors; no order |
| cast | Categorical | Names of actors; multiple per entry |
| country | Categorical | Country names; no order |
| date\_added | Interval | Date information, useful for time-based trends |
| release\_year | Interval | Year of release; no true zero, so it's interval |
| rating | Ordinal | Content rating (e.g., G < PG < PG-13 < R < TV-MA) |
| duration | Ratio | Time in minutes or number of seasons; has a true zero |
| listed\_in | Categorical | Genre(s); one or more categories |
| description | Categorical | Text description; not numeric or ordered |

**4. Domain Specification for Each Column:**

The domain of a column refers to the range or list of values that the column can take. Below is the domain for each column in the Netflix dataset:

1. **show\_id:** This is a unique identifier for each entry in the dataset. It contains 8,807 unique values such as s1, s2, s3, and so on.
2. **type:** Indicates whether the content is a Movie or TV Show. The domain includes two values: Movie, TV Show.
3. **title:** Contains the name of the show or movie. There are 8,807 unique titles representing individual content.
4. **director:** Names of directors for the content. It includes around 4,528 unique values. Some entries may be missing (unknown directors).
5. **cast:** Lists the actors/actresses involved in the content. There are 7,692 unique cast combinations. This field may also be empty in some cases.
6. **country:** Specifies the country or countries where the content was produced. The domain includes 748 unique country names, with some entries listing multiple countries.
7. **date\_added:** The date the content was added to Netflix. There are 1,767 unique dates, formatted as "Month Day, Year" (e.g., September 25, 2021).
8. **release\_year:** The year when the content was originally released. The domain ranges from 1925 to 2021.
9. **rating:** This column contains 17 different content ratings used to indicate suitability for viewers (e.g., PG, TV-MA, R, G, NC-17, etc.).
10. **duration:** Represents either the duration of a movie (in minutes) or the number of seasons for a TV show. There are 220 unique values like 90 min, 1 Season, 2 Seasons, etc.
11. **listed\_in:** Lists genres or categories associated with each title. There are 514 unique combinations (e.g., Comedies, International TV Shows, Dramas).
12. **description:** Provides a summary of the show or movie. The domain includes 8,775 unique text descriptions, with some entries missing.

**5.** **Target Users for the Netflix Dataset Dashboard**

A dashboard built using the Netflix Movies and TV Shows dataset would be valuable to several types of users in the **entertainment and media industry**, as well as adjacent fields that deal with data analytics, content strategy, and consumer behavior.

### **Primary Users:**

* **Content Strategists at Streaming Platforms**: They can analyze trends in content types, genres, and release years to identify what kind of content is in demand and which categories are underrepresented.
* **Entertainment Data Analysts**: These users would benefit from visualizing metrics like the rise of international content, popularity of TV shows vs. movies, or shifts in genre trends over time.
* **Marketing and Audience Engagement Teams**: By understanding content distribution across countries, ratings, and genres, marketing professionals can tailor campaigns to different audience segments more effectively.
* **Production Companies & Studios**: They can use the dashboard to study competitor content strategies, identify popular themes, and spot market gaps for new productions.
* **Academic Researchers and Media Students**: Useful for analyzing global media trends, cultural representation, and shifts in public viewing preferences based on rating and genre patterns.

**6**. **Prospective Users and Their Use Cases**

We will develop the Netflix content dashboard primarily for the following user groups:

1. **Content Strategists at Streaming Platforms**: To identify trends in genres, popular content types (movies vs. TV shows), and audience preferences based on ratings or regions. This helps in planning what kind of shows or movies to license or produce next.
2. **Marketing and Audience Targeting Teams**: To analyze which content appeals to different demographics and countries. This insight allows them to run more personalized and effective marketing campaigns.
3. **Entertainment Industry Analysts**: To study historical trends in content production, global expansion of Netflix, and genre evolution over time. Useful for industry reports and market insights.
4. **Media and Film Students / Academic Researchers**: To explore how content distribution varies across genres, regions, and time. The dashboard can serve as a learning or research tool for studying global streaming trends.
5. **Production Companies and Studios**: To examine what types of content are missing or oversaturated in Netflix’s catalog. This can help them pitch shows that fill content gaps or align with popular trends.

**7**. **Comprehensive Set of User Questions**

1. **How many movies vs. TV shows are currently available on Netflix?**
2. **What are the most popular content ratings (e.g., TV-MA, PG-13, R) on Netflix?**
3. **Which countries produce the most content available on Netflix?**
4. **How has the number of new titles added to Netflix changed over time (by year or month)?**
5. **What are the most common genres or genre combinations across all titles?**
6. **Which years had the highest number of releases?**
7. **What is the average duration of movies on Netflix? Are there any noticeable trends by year or rating?**
8. **Which directors have the most titles featured on Netflix?**
9. **Which actors or actresses appear most frequently across Netflix content?**
10. **Are there genre preferences by country (e.g., India vs. US vs. Japan)?**
11. **How are titles distributed across different maturity ratings by type (movie vs. TV show)?**
12. **What are the top 10 most represented countries in each genre (e.g., drama, comedy, horror)?**
13. **Are more international titles being added to Netflix in recent years compared to earlier years?**
14. **How many one-season vs. multi-season TV shows are available?**
15. **What genres are most common among TV shows vs. movies?**

**References**

**1. Netflix Movies and shows Dataset:** <https://www.kaggle.com/datasets/shivamb/netflix-shows>

**2. Mural link:** https://app.mural.co/t/ift5338876/m/ift5338876/1744948618212/1841962e5e3801a8ff05bc1b200746c861dcb20b?sender=u028cea11dd94b51431478940