**DATA VISUALIZATION (ISM6419.901F23)**

**FINAL PROJECT REPORT**

**Analyzing Netflix movies and TV shows Using IMDB**

**Name: Bandi Rohan Reddy**

**Introduction:**

The way we consume media has changed dramatically because of streaming services like Netflix, moving away from traditional broadcast formats and toward on-demand digital material. To examine Netflix's library of films and TV series, this project makes use of IMDb, a premier source of information on movies and television. For this research, IMDb ratings are a useful statistic since they act as a stand-in for popularity and content quality. This initiative attempts to identify trends and insights that signify success within the streaming platform's user base by merging IMDb's ratings with Netflix's vast data.

**Ambitiousness of the Project:** This project has a broad scope as it aims to explore the huge quantity of content data to determine what motivates Netflix users to participate and feel satisfied. It aims to bridge the gap between numerical ratings and qualitative success, considering various factors such as genre, directorial influence, and trends over time. With this research, a sophisticated understanding of the content strategy that may influence the direction of streaming services in the future may be obtained.

Moreover, this project makes sense of Netflix's movie and TV show data by creating simple, clear visuals. We use these visuals to show how IMDb ratings connect to what people watch on Netflix. It's like drawing a map that points out the popular spots and the hidden gems, making it easy for everyone to see the big picture.

**Further Research Questions:**

**1.Content Lifespan and Trends:**

What content characteristics from IMDb data are indicative of a longer lifespan on Netflix and potential for becoming evergreen titles?

**2.Cross-Platform Performance Analysis:**

How does content performance on Netflix, as reflected by IMDb ratings, compare to its performance on other streaming platforms?

**3.Impact of Critical Acclaim on Subscriber Growth:**

Does critical acclaim on IMDb, such as high ratings and awards, translate to subscriber growth and retention for Netflix?

**4.Machine Learning for Content Success Prediction:**

Can machine learning models using IMDb data accurately predict the success of future Netflix content in various regions and demographics?

**Methodology:**

In this project, I used five datasets that covered various aspects of Netflix TV Shows and Movies and IMDB These datasets were sourced from **Kaggle** (<https://www.kaggle.com/datasets/thedevastator/the-ultimate-netflix-tv-shows-and-movies-dataset>) and IMDB (<https://www.kaggle.com/code/adrianmcmahon/imdb-indian-movies-eda/input>) and Statista https://www.statista.com/statistics/583249/netflix-series-viewership/

**Dataset1:netflix\_titles  
Attributes**: type, title, director, description, Listed\_in, duration, cast, country, Date\_added, Release\_year

**Dataset2:** **netflix\_IMDB**

**Attributes**: type, title, director, cast, country, date\_added, release\_year, rating, duration, listed\_in, description, titleType, primaryTitle, originalTitle, isAdult, startYear, endYear, runtimeMinutes,genres, averageRating, numVotes, year\_added, month\_added, day\_added, country\_count

**Dataset3: IMDB Movies India**

**Attributes**:Name, Year, Duration, Genre, Rating, Votes, Director, Actor 1

**Dataset4: Best Movies Netflix**

**Attributes**:TITLE, RELEASE\_YEAR, Ratings, NUMBER\_OF\_VOTES, DURATION, MAIN\_GENRE, MAIN\_PRODUCTION

**Dataset5:** **Best Shows Netflix**

**Attributes**:TITLE, RELEASE\_YEAR, Ratings, NUMBER\_OF\_VOTES, DURATION, NUMBER\_OF\_VOTES, DURATION, NUMBER\_OF\_SEASONS, MAIN\_GENRE, MAIN\_PRODUCTION

**Connection:**

A white text on a white background

Description automatically generated

**Analysis:**

**Q1. How does the average rating of titles directed by various filmmakers correlate with their popularity index on Netflix, and does this relationship differ between movies and TV shows?"**

A screen shot of a graph

Description automatically generated

The data visualization shown demonstrates the difference between the average rating of filmmakers on Netflix and their level of popularity. The popularity metrics of directors differ greatly, even though most of them have average ratings that are about 6.3 to 6.4. Despite having a comparable average rating, Aaron Sorkin stands out with a popularity index of 456,525, far exceeding others. Directors such as Abdelaziz and Aaron Burns, on the other hand, have minor popularity indices of 9 and 10, respectively. This significant variance implies that a filmmaker's average rating is not directly correlated with their level of popularity on Netflix.

I used a calculated field for this, named Popularity Index. **(([Num Votes] \* [Average Rating]) / 1000)**

**Q2. Which country has the highest release on Netflix?**

A screenshot of a map

Description automatically generated

The map uses a color gradient to depict the volume of content, with darker shades representing higher counts. The United States has the highest count with 1,759 releases, followed by India with 644, and the United Kingdom with 193. Other countries with notable counts include Canada (93), Brazil (68), Australia (132), and Russia (68). Smaller counts are visible throughout Europe, Asia, and Latin America, indicating a diverse but less concentrated availability of Netflix content.

**Q3. Compare the difference between top 10 different genres?**

**A screen shot of a diagram

Description automatically generated**

The bubble chart in the image illustrates a comparison between different Top 10 genres, with each bubble's size representing the number of titles. Documentaries are the most abundant genre with 359 titles, slightly surpassing Stand-Up Comedy at 334 titles. Dramas with an international focus are prominently featured, with one category having 362 titles and another comprising Dramas and International Movies with 180 titles. The Children & Family segment is well-represented with 215 titles exclusively and 201 titles when combined with Comedies. Additionally, genres that blend Comedies, Dramas, and International Movies account for 274 titles, indicating a robust selection in cross-genre offerings. The visual representation of Netflix's most popular genres. In addition to showing possible areas of concentration for content generation or acquisition, this graphic may clarify how different types of material are distributed around the platform.

**Q4.** **What trends can be observed in the IMDB ratings of Indian movies over the years?**

**A graph with blue dots and black text

Description automatically generated**

The scatter plot from the Tableau illustrates the average IMDb ratings for Indian movies on Netflix over a historical timeline. Each point reflects the average rating for movies released each year, with the trend indicating a slight but statistically significant decline in ratings over time. The regression line, defined by the equation Avg. Rating1 = -0.0108827\*Year + 27.5732, has a negative slope, suggesting ratings have decreased on average by 0.01088 points per year. Despite this, the R-squared value is 0.190154, meaning that only 19% of the variability in the ratings is explained by the year of release, highlighting that other factors are influencing these ratings. The data spans from the early 20th century to the present, with ratings varying between approximately 5 and 8.5, indicating a moderate to high average quality of Indian films on Netflix according to IMDb ratings. The p-value of less than 0.0001 confirms the significance of the trend, even though the effect of time on ratings is quite small.

**Q5. What were the best movies and TV shows released in each year?**

A screen shot of a graph

Description automatically generated

The dual-axis bar and line chart illustrates the distribution and trend of high-quality content from around 1950 to 2022. The chart indicates a gradual increase in the number of movies over the years, with a more pronounced rise from the 2000s onward. The blue bars, representing movies, show fluctuations in count but no clear pattern, suggesting variable acquisition or production rates. In contrast, the orange line, representing TV shows, demonstrates a relatively stable count until it surges around 2022, where the line peaks sharply. This peak represents the highest count of TV shows released in a single year, as recorded on the chart.

**Q6.** **How the release years and average ratings of movies and TV shows on Netflix evolved, and what patterns emerge from the data in terms of content growth and quality over the years?**

A graph of growth and statistics

Description automatically generated with medium confidence

Dual-axis chart reveals a marked disparity in the growth trajectories of Netflix's movie and TV show libraries. From 1990 to 2021, the movie library experienced a dramatic expansion, peaking in 2020 with 23,944 titles, before a sharp downturn occurs in 2021. TV shows, on the other hand, exhibit a steady increase to 315 titles by 2021, without a subsequent decline. In terms of ratings, movies display a slight downward trend in average IMDb ratings over the years, while TV shows show a relatively stable and marginally increasing trend in ratings.

**Q7. How does viewer engagement, as measured by the number of votes, vary with the duration (number of seasons) of TV shows across different genres on Netflix?**

**A screen shot of a graph

Description automatically generated**

The stacked bar chart illustrates Netflix show longevity and viewer engagement, measured by duration (number of seasons), and sum of IMDb votes, respectively. Genre-specific engagement is depicted by color, with some genres demonstrating significantly more votes, suggesting higher engagement.

**A screen shot of a graph

Description automatically generated**

Both charts depict engagement through the volume of votes, revealing which genres garner more sustained interest and potentially influence the duration of a show's run on the platform. This data can inform content strategy, indicating which genres might be more viable for long-term investment based on viewer engagement levels.

**Q8.** **How does the availability of Netflix content across countries compare between movies and TV shows over the years, and what does the trend indicate about Netflix's global distribution strategy?**

**A graph of a line graph

Description automatically generated with medium confidence**

The line graph displays a steady increase in the country availability of Netflix movies, peaking at 3,891 countries in 2021, while TV shows peak at 315 countries around 2019 before a sharp decline. The data indicates a broader international distribution for movies over TV shows and suggests a recent contraction in TV show availability.

**Q9. How do the average ratings for movies and TV shows across different genres on Netflix compare, and which genre receives the highest ratings in each category?**

A screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

The first chart displays average ratings for best Netflix movies and TV shows across various genres. The second chart applies a filter for "Best Movies and TV Shows Netflix " showing a separate set of average ratings for top Movies and TV Shows within those genres. These visualizations help identify which genres have higher average ratings and how the top movies and TV Shows influence those ratings.

**Q10.** **Which actors are most frequently associated with high-rated content on Netflix, indicating a potential "star power" effect?**

**A graph of different colored lines

Description automatically generated with medium confidence**

The image shows an area chart detailing the frequency of high-rated versus not high-rated movies across a range of actors, labeled as "Actor 1" . The areas are color-coded with orange representing movies that are not high-rated and blue representing high-rated movies. The chart suggests variability in the count of high-rated films for each actor, with some having a substantially larger area of blue, indicating a higher frequency of high-rated movies in their filmography.

I used a calculated field for this, named High Rated **(IF [Average Rating] >= 7 THEN "High Rated" ELSE "Not High Rated" END)**

**Conclusion:**

After analyzing the Visualizations, we can conclude that Netflix has a flexible and customized content strategy., with certain directors and actors consistently influencing high ratings and viewer engagement, indicative of a "star power". Genre popularity varies, with trends reflecting changing viewer preferences and market targeting strategies. The distribution of content reveals a strategic global expansion, with particular attention to regional and content performance. Finally, the longevity and popularity of TV shows on Netflix appear to be affected by the number of seasons, highlighting the importance of sustained narrative engagement in viewer retention.