**Final Project Report – ISM6419**

**Data Visualization**

**Spring 2023**

**Netflix Content Data Analysis**

**USF MUMA COLLEGE OF BUSINESS, Tampa**

**Rajashekar Reddy Patlori(U57791061)**

**Introduction:**

Netflix is a streaming service that has changed the way people watch movies. Netflix's impact in daily life extends beyond mere entertainment. It provides many individuals with opportunities for education, inspiration, and cultural interaction. The website provides a varied selection of information from across the world, allowing users to learn about various cultures, languages, and points of view. Exposure to new ideas and experiences may be enlightening and transformational, expanding perspectives and challenging assumptions.

Reference: <https://en.wikipedia.org/wiki/Netflix>

**Ambitiousness of the project:**

Given the size and complexity of the data involved, the Netflix data visualization project is an extremely ambitious endeavor. The effort entails gathering and analyzing data from Netflix's massive library of material. What makes this project especially challenging is the sheer volume and variety of data that must be analyzed. Netflix boasts millions of continually growing collections of content that are added and removed on a regular basis. The project will require advanced data analytics tools and techniques, as well as a thorough understanding of content trends, to make sense of this data**.**

**Research Questions**

Question 1:

What is the trend in distribution and diversity of Netflix content based on production countries and number of shows and movies availability over the years?

Question 2:

What are the most popular genres, most prolific and successful directors and most featured actors in movies and TV series on Netflix?

Question 3:

How has the age certification of Netflix content evolved over time, and what are the trends in the types of content produced for different age groups?

Question 4:

How has the addition of new movies and TV shows to Netflix varied monthly from 2015 to 2021, and are there any seasonal or yearly trends in this variation? Furthermore, is there a pattern in the weekday releases of new content on Netflix, with certain weekdays being more popular than others?

Question 5:

Is there a relationship between the awards received by movies and their box office collections, and if so, what is the nature of this relationship?

**Methodology**

To get the required insights, it is important to get the required data from different sources and do the preprocessing to make sure it is not biased.

The data is sourced from:

[https://flixgem.com](https://flixgem.com/)

<https://www.kaggle.com/datasets/shivamb/netflix-shows?resource=download>

<https://www.imdb.com/interfaces/>

<https://www.rottentomatoes.com/>

<https://www.metacritic.com/>

<https://data.world/chasewillden/netflix-shows>

Web scraping was utilized to construct the data files, and I obtained some straight data columns from Kaggle and Data World.

Many missing fields, duplicates, and NULL values were added by getting the details from <https://www.netflix.com/> and some records for which I couldn't find the details were dropped to sync all the data sources..

**Analysis**

The dataset has been put to the tableau, and links based on the Show titles and Show Ids have been built to combine all the data.

**Research Question 1:**

**What is the trend in distribution and diversity of Netflix content based on production countries and number of shows and movies availability over the years?**

For the following research question, I have constructed 4 different visualizations.

**A picture containing circle, screenshot, diagram, graphics

Description automatically generated**

1.1 Movies to Shows Distribution Ratio

* The visualization provides the understanding of the movie to series ratio in Netflix app.
* From the pie chart we get the insight that around 71.5% of the content in Netflix is movies and the rest 28.5% are series.
* So, we can conclude that content in Netflix is dominated by movies.

***A picture containing text, plot, screenshot, diagram

Description automatically generated***1.2 Number of Shows and Movies on Netflix by Year

* The following visualization shows the distribution of movies to series over the years.
* From this Area graph we can see that after 2015 the addition of content to Netflix application has drastically increased.

A map of the world

Description automatically generated

1.3 A Visual Exploration of Netflix Content by Production Country

* The graph above helps us realize how diversified the content on Netflix is. The movies and TV Series on Netflix are produced in almost every country in the globe.
* We can see the most content is produced in USA and followed by INDIA.

A picture containing screenshot, text, diagram

Description automatically generated

1.4.1 Top Countries by Movies Produced

A picture containing screenshot, text, diagram

Description automatically generated

1.4.2 Top Countries by Series Produced

* The above Tree graphs show the top Netflix content producing countries.
* The topmost most Tree graph indicates that top movie producing countries are United States of America, India, United Kingdom, Japan, and South Korea
* The second Tree graph indicates that top tv series producing countries are United States of America, Japan, South Korea, United Kingdom, and India

**A picture containing text, map, diagram, screenshot

Description automatically generated**

1.5 Distribution Dashboard

The following dashboard gives us insights to research question 1 where it analyzes the trend in distribution and diversity of Netflix content based on production countries and number of shows and movies availability over the years.

**Research Questions 2:**

What are the most popular genres, most prolific and successful directors and most featured actors in movies and TV series on Netflix?

The following three visualizations are going to show the details.

A screenshot of a graph

Description automatically generated with medium confidence

2.1.1 Top Genres in Movies

A screenshot of a graph

Description automatically generated with medium confidence

2.1.2 Top Genres in TV Series

* The above bar graph shows us the top Genres in Movies and TV Series
* The top genre in movies is comedy followed by documentation.
* The top genre in TV Series is drama followed by documentation.

**A picture containing text, circle, design

Description automatically generated**

2.2 Top Directors by Movies Directed

* The above graph shows us the top directors by number of movies in Netflix.
* The top directors by number of movies in Netflix are Marcus Raby, Jay Karas, [Raul Campos, Jan Suter], Cathy Garcia-Molina and Jay Chapman.

**A picture containing screenshot, text, line, font

Description automatically generated**

2.3.1 Actors featured in most Movies.

**A picture containing screenshot, text, line, font

Description automatically generated**

2.3.2 Actors featured in most Series.

* The above graph shows us the most featured actors by number of movies and series in Netflix.
* The most featured actor in movies is Shah Rukh Khan.
* The most featured actor in the series is Takahiro Sakurai.

**A screenshot of a computer

Description automatically generated with medium confidence**

The following dashboard gives us insights to research question 2 where it explains about the most popular genres, most prolific and successful directors and most featured actors in movies and TV series on Netflix.

**Research Question 3:**

How has the age certification of Netflix content evolved over time, and what are the trends in the types of content produced for different age groups?

A picture containing text, screenshot, rectangle, diagram

Description automatically generated

3.1 Evolution of Netflix Content by Age Certification and Year

* The above graph explains about the distribution of content by Age Certification.
* So, the above graph concludes that for the last 5 years most of the content in Netflix is TV-MA ratings which means it’s for mature audiences.

**Research Question 4:**

How has the addition of new movies and TV shows to Netflix varied monthly from 2015 to 2021, and are there any seasonal or yearly trends in this variation? Furthermore, is there a pattern in the weekday releases of new content on Netflix, with certain weekdays being more popular than others?

A picture containing text, handwriting, font, line

Description automatically generated

4.1.1 Number of Movies Added by Each Month [2015-2021]

A picture containing text, handwriting, font, line

Description automatically generated

4.1.2 Number of Series Added by Each Month [2015-2021]

* The above animated line graph explains about the number of Movies and Series added per each month from 2015 to 2021.
* There were few important patterns in this animation the content is Netflix is increasing every year.
* And in and after 2019 Netflix increased the rate of adding TV series when compared to TV Series this is right after the Covid Pandemic
* And in the months of Christmas break and Summer break the content added to Netflix was more when compared to other months

A picture containing text, screenshot, diagram, circle

Description automatically generated

4.2.1 Movies Release Proportion by Weekdays

A picture containing diagram, screenshot, text, circle

Description automatically generated

4.2.2 TV Series Release Proportion by Weekdays

* The above pie graph explains the content release per weekday ratio.
* We can clearly see that most of content is released on Fridays be it Movies or TV series.

**Research Question 5:**

Is there a relationship between the awards received by movies and their box office collections, and awards received to award nominations and if so, what is the nature of this relationship?

A picture containing parallel, line

Description automatically generated

5.1 Awards to Collections Relationship in Movies

* The above dual axis bar to line graph depicts the relationship between awards collection to box office collections of movies in Netflix.
* The graph explains that there is a relationship between the fields are corelative, mostly the movies with high box office collections bagged awards.

**A picture containing text, screenshot, line, plot

Description automatically generated**

5.2 Movies with Award Nominations to Awards Winnings

* The above scatter depicts the relationship between awards collection to awards nominations of movies in Netflix.
* The trend line in the graph indicates that awards nominations are directly proportional to awards received.
* Mostly all the movies followed this trend but still there are few outliers.

**Few Exploratory Visualizations**

A picture containing text, screenshot, logo, design

Description automatically generated

* The above dashboard shows the top-rated movies and shows in various rating websites.

**A screenshot of a computer

Description automatically generated with medium confidence**

* The above box plot depicts shows the top-rated movies and shows in various rating websites.
* Most of the movies have a runtime of 1-2 hrs.

**CONCLUSION**

This project will assist us in understanding the results of the research questions mentioned above, as well as the various insights and patterns that will assist us in understanding the growth of the OTT industry, the distribution and diversity of content around the world, and how it is distributed among various age groups.

The details are mentioned below:

* There is a clear analysis of the trend in distribution and diversity of Netflix content.
* Trends in growth of content over time and yearly and monthly patterns in release of content
* The most added content int Netflix are movies; USA is the most content produced country.
* The content added to Netflix is mostly PG-MA ratings technically its suitable to audience only above 17 years.
* The most viewed genre in movies and series are comedy, drama respectively while the second most viewed genre is documentation.
* Analyzed correlation between independent fields such as Box office collections and Awards bagged.
* Content addition trends in Netflix from the analysis:
  + During and post covid the rate of shows/series added are extremely high
  + It is observed that during winter & summer breaks the content added is relatively more than other months.
  + Friday is the entertainment day for the crowd hence most content is added on this day of the week.

**FUTURE RESEARCH QUESTIONS:**

* What insights can be gained from analyzing user ratings and reviews of movies and TV shows on Netflix, and how can these insights be used to make personalized recommendations for individual users?
* What is the relationship between Netflix's stock price and its user subscription numbers, and how do changes in user subscriptions impact the company's financial performance and stock market valuation?