

# **Decoding OTT: Uncovering global insights for film industry advancement and cultural exchange**

**Submitted for**

**DATA VISUALIZATION AND DASHBOARD**

**Submitted by:**

**Vidisha Arvind (E23MTCG0004)**

**Abhay Kumar (E23MCAG0134)**

**Ayush Kumar (E23MCAG0132)**

**Submitted to-**

**Dr. Nitin Arvind Shelke**

**July-Dec 2023**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**



## INDEX

---

Sr No	Content
1.	<b>ABSTRACT</b>
2.	<b>INTRODUCTION</b>
3.	<b>RELATED WORK</b>
4.	<b>TOOLS &amp; IDE USED</b>
5.	<b>METHODOLOGY</b>
6.	<b>RESULTS</b>
7.	<b>CONCLUSION</b>
8.	<b>REFERENCES</b>

## LIST OF FIGURES

Figure No.	Title
Fig. 1	No Outlier Observed in the above boxplot
Fig. 2	Brand color palettes for Disney, Netflix, and Amazon
Fig. 3	Block Diagram
Fig. 4	Distribution of content kinds
Fig. 5	Heatmaps of Amazon prime, Hotstar and Netflix indicating age groups associated with particular OTT
Fig. 6	Bar Graph Showcasing the topmost 10 countries in content production

## **ABSTRACT**

This research aims to decode the impact of Over the Top (OTT) platforms on the worldwide film industries and cultural exchange. By finding insights from data of top OTT platforms such as Amazon Prime, Hotstar Disney plus,Netflix etc. insights into audience preferences, market trends, and cultural interaction through films are obtained. The study proposes techniques to improve user experience, including predicting ratings and trends, and clustering similar content based on features of the content like genre, titles etc. The findings contribute to better content production, personalized recommendations, and market expansion in the OTT industry.

## **INTRODUCTION**

The landscape of streaming platforms i.e. OTT, which emerged in the early 2000s, witnessed its inception in India around 2008. Notably, industry giant Netflix made its foray into the Indian market in January 2016. As of today, OTT platforms have undeniably wielded a substantial influence on the entertainment industry, offering a plethora of captivating content. However, amidst the commendable progress, significant challenges have emerged, including hurdles in content discovery, content restriction and regulatory compliance. There is a struggle to strike a balance between artistic freedom and cultural and legal standards which is still going on as these platforms offering wide variety of content, including documentaries, TV shows, and movies. The complex terrain in question often gives rise to disputes regarding content appropriateness, which may impede the creative expression of filmmakers and content suppliers.

The emergence of Over the Top (OTT) streaming services have changed the entertainment industry by giving consumers a substitute method of consuming material. Over-the-top (OTT) services, like Netflix and Prime Video, offer a wide range of movies, TV shows, and original programming that can be seen whenever and wherever you choose. With a rise in subscriptions and a significant amount of user engagement, the over-the-top (OTT) industry has been becoming popular in

India in recent times. With the arrival of over-the-top (OTT) platforms in India, users now have access to an extensive library of content that suits a wide range of tastes and inclinations. But in addition to the advantages, there are some drawbacks that must be taken into consideration.

For addressing some major challenges with content streaming platforms, our study endeavors to devise comprehensive solutions that mutually benefit both the entertainment businesses and users alike. By delving into the intricacies of content curation, recommendation algorithms, and user feedback mechanisms, We want to improve the entire OTT experience. Our motive is to foster an environment where content discovery becomes seamless, preferences are duly acknowledged without bias, and user satisfaction is paramount. Through this research-oriented approach, we aspire to pave the way for a more harmonious relationship between OTT platforms and their discerning audience.

One of the major key issues faced by OTT platforms is biased content recommendation. The recommendation algorithms used by these platforms often fail to provide personalized and relevant content suggestions to users. This leads to frustration among users who face issues in discovering new content that aligns with their interests. As a result, users may become dissatisfied and ultimately churn from the platform. Churn, in the context of OTT platforms, refers to the phenomenon where users cancel their subscriptions or stop engaging with the platform. The churn problem is a significant concern for OTT platforms as it directly impacts their revenue and growth. Poor content discovery and biased recommendations contribute to customer frustration, leading to higher churn rates.

## **RELATED WORK**

A. NLP based recommendation system for sentiment polarity.

In their study Sachin Bohite, C.H. Patil , Saurabh Pal, Vikas Magar have sought to infer a new sentiment-based system for recommending films. During our study, we discovered that machine learning suggestions based on polarity attitudes are more

powerful than recommendations based on categories or ratings, which can be executed with any of the machine learning approaches and have approached Python to create our RS. Our RS is built on the KNN technique and cosine similarities. The use of vectorization to compare emotional data. If the vectors are similar, the algorithm suggests movies that are comparable[8].

B. Movies on Ott study in R using random forest techniques and multiple regression methodologies.

As pointed out by D. Patil, movies are a global source of entertainment as well as an effective medium for educating or indoctrinating individuals. In light of the present pandemic crisis, for people all across the world, one of the most interesting and important ways to relax is by watching content on streaming applications. In order to gather insightful information, this research will cover every film that exists on widespread streaming services, such as OTT. A Kaggle data set that was assembled from the following streaming services: Netflix, Prime Video, Hulu, and Disney's streaming service is used to do this. All of the films, their evaluations, and the corresponding streaming websites where they may be found are included in the dataset[3]. It provides comprehensive details about every movie, including the writer and director, language, genre, year of launch, and IMDB rating.

C. A Study of Multimedia Service Experience Qualities and Machine Learning-Based Forecasting

Vladimir Poulkov, Pavlos, Georgios Kougioumtzidis, Zaharias D. Zaharis, and Lazaridis In their survey, they underlined the need of machine learning (ML) based QoE (Quality of experience) prediction models for QoE management in services involving multimedia. The study examines novel approaches and difficulties connected to evaluating the quality of multimedia services, with a focus on extended reality and video gaming applications, and provides an overview of cutting-edge ML-based QoE prediction models. Furthermore, the study defines QoE in the context of multimedia services and provides a comprehensive analysis

of QoE influencing variables. In addition, the survey collects and analyzes the most significant subjective and objective quality indicators, as well as methods for analyzing their performance and mathematical models for linking QoS parameters with QoE[9].

#### D. M5: Multi Scenario Matching for Over the Top Recommendation through Multi-Modal Multi-Interest

Xin Gao, Pengyu Zhao, Liang Chen, Chunxu Xu, and Xin Gao used multi modal multi scenario multi interest matching (M5) in order to take advantage of the unique characteristics of the OTT recommendation. To gain access to the rich metadata information, M5 suggests a multi-modal incorporating layer. To capture users' various interests, it offers a multi-interest extraction layer. To facilitate effective information transformation across multiple scenarios, it offers a multi-scenario mixing layer. Finally, it offers a weighted candidate matching layer to merge the hybrid user-item preferences[10]. Thorough investigation of the Hulu platform and Disney Plus Hotstar demonstrates M5's excellence in industrial OTT services, emphasizing the use of task-specific characteristics in model creation.

#### E. Evaluation of Machine Learning (ML) Utilization in OTT platform recommendations

A. Rao and A. Khurre[11] have demonstrated how to integrate machine learning techniques in various OTT platform recommendation systems to enhance the platform's users' experience. Furthermore, this study targeted at describing several types of RS and discovered that content based filtering is not as effective as collaborative filtering, and that the accuracy of any recommender system may be enhanced by integrating additional movie or a content form elements. Based on this research, we can infer that hybrid filtering is more successful in determining user preferences across various OTT platforms. It contributes to the development of an effective and efficient recommendation system by integrating the skills of the other two techniques. As a result, we became aware of the possibility to enhance the OTT

platform's recommendation algorithms through the application of various Machine Learning techniques.

F. The variables influencing the Over-the-Top (OTT) streaming platforms that consumers select.

According to the research of Shalawat and colleagues, the OTT media business continues to expand quickly, causing the need for the creation of plans to bring in and attract even more customers. The majority of individuals view material on platforms for longer periods of time, with subscriptions increasing during and after lockdowns. Integrated marketing communication and influencing consumer choices for streaming OTT platforms are critical for increasing subscribers and customer satisfaction.[1]

G. The Emergence of Over-The-Top Platforms: Shifting Customer Preferences

This study aims to predict the needs of future generations and the elements that led to Indian consumers' acceptance of online streaming services. The systems will be able to control greater loads since the network will be more competent. Applications in great demand indicate simultaneous HD video streaming and virtual reality experiences. Because of the steady rise in internet usage, OTT platforms have a bright future. This report emphasizes how the arrival of the Covid pandemic has put the world in a terrible condition. The Coronavirus has impacted a great number of enterprises, such as those in the travel and tourist sector, the OTT platform, and other firms involved in online content commerce. [7]

H. Analyzing consumer preferences and patterns for OTT platforms:

According to R. S. Dixit et al., to improve movie and series viewing experiences, the media industry introduced over-the-Top (OTT) platforms[2]. These subscription-based services, which offer benefits such as lower costs, high-quality content, and repetitive playback, are replacing cable TV, cinema, and DTH. The purpose of this study is to learn about consumer preferences for OTT platforms in



Jaipur. The study analyzes demographic factors, traditional and new entertainment preferences, and the relationship between OTT platform usage and subscribed platforms using multiple regression tools and R software. The findings show no statistically significant difference between traditional and new entertainment spending. Furthermore, the study found that different occupational groups prefer different OTT platforms. Overall, OTT platforms provide consumers with a more convenient and cost-effective alternative to traditional television and movie theater services.

#### I. An algorithm utilizing machine learning to select the most effective Ott platform.

The main aim of this study is to investigate method for multiclass classification issues and how to reflect them to the process of choosing an productive over-the-top Platform Investigate Techniques include reinforcement-based systems, supervised learning, and unsupervised learning, taking into account the function of cognitive technologies in processing several forms of data. An algorithm admit for its efficacy in other fields, such as bioinformatics, is used to pick out the OTT platform. The study shows that adopting Random Forest to determine which over-the-top (OTT) platform is the most effective is a feasible approach when taking into account variables like content range, video streaming quality, and subscriber base.

#### J. Over-the-top (OTT) video services' emergence and potential in India

N. Elangovan and E.Sundaravel have stated that Hotstar is currently leading the video streaming sector in India, where it has become a successful content consumption strategy. The most frequently utilized gadgets for streaming OTT video content are smartphones, with Xiaomi being the most well-known manufacturer. Traditional TV stations should concentrate on producing high-quality content and get ready for the paradigm shift brought about by OTT platforms in order to compete with OTT content[14]. The Indian internet gaming sector is anticipated to reach billions of dollars in valuation by 2020. Television continues to

be the largest subsegment of the entertainment and media sectors, despite increased video use. The CAGR for the Indian television sector is predicted to be 10.6% in 2022.

#### K. The Dynamics of OTT Platforms for Modern Indian Theatre

According to this study, it is often recognised that OTT affects theater production and the creative process. The notion that OTT will supplant theater has started to terrify some in the theater business. In the future, the theater board should provide authors and directors with further guidelines for what they can portray[15] varied subjects that need information. Since India has insufficient screening, there ought to be a greater number of displays. The rapid development of over-the-top (OTT) platforms as a preferred global entertainment channel poses a severe threat to traditional forms of entertainment.

#### L. Unstoppable Over-The-Top - An Overview of OTT Platform and Regulation in India

The OTT content industry is still quite young universally. Over-The-Top (OTT) service customers have increased remarkably in the majority of countries. Certain nations have ratified laws or are thinking about enacting laws to control the content that is obtainable on over-the-top (OTT) services. The legislation appears to be required in order to address definite gaps that the country is working to close. Owing to the unexpected shift in the entertainment industry to the internet, consumers have access to a wide field of information. Spectators are drawn to a variety of creative and genre content due to the bounty of possibilities offered by over-the-top (OTT) services[16]. Therefore, governments must take into account the interests of their citizens when they gallop to pass legislation regulating material.

### Summarization of related work

S.no	Paper title	Dataset used	Methodology	Results
1	NLP-based Recommendation System for Sentiment Polarity	Not specified	KNN technique and cosine similarities	Machine learning suggestions based on polarity attitudes are more powerful than category or rating-based recommendations.
2	Movies on OTT Study in R using Random Forest Techniques and Multiple Regression Methodologies	Kaggle dataset from Netflix, Hulu, amazon prime	Dataset analysis	Insights about different platforms
3	A Study of Multimedia Service Experience Qualities and Machine Learning-Based Forecasting	Not specified	Machine learning based QoE prediction models	Examines novel approaches and difficulties in evaluating multimedia service quality
4	M5: Multi-Scenario Matching for Over-the-Top Recommendation through Multi-Modal Multi-Interest	Not specified	Multi-modal multi-scenario multi-interest matching (M5)	M5's excellence in industrial OTT services

5	Evaluation of Machine Learning (ML) Utilization in OTT Platform Recommendations	Not specified	Integration of machine learning techniques in various OTT platform recommendation systems	Hybrid filtering is more successful in determining user preferences across various OTT platforms.
6	Variables Influencing Over-the-Top (OTT) Streaming Platforms that Consumers Select	Survey Data	Surveyed individuals	Integrated marketing communication and influencing consumer choices are critical for increasing subscribers and customer satisfaction.
7	The Emergence of Over-The-Top Platforms: Shifting Customer Preferences	Not specified	Customer preferences were analyzed.	Highlights factors leading to Indian consumers' acceptance of online streaming services and the impact of the COVID-19 pandemic.
8	Analyzing Consumer Preferences and Patterns for OTT Platforms	Not specified	Multiple regression tools and R software	OTT platforms provide consumers with a more convenient and cost-effective alternative to traditional television and movie theater services.

9	An Algorithm Utilizing Machine Learning to Select the Most Effective OTT Platform	Not specified	Reinforcement-based systems were utilized.	Random forest outperformed for the analysis of Ott data to understand trends.
10	Over-the-Top (OTT) Video Services' Emergence and Potential in India	Kaggle dataset from Hotstar	Analysis of whatever devices used for streaming of Ott content.	The CAGR for the Indian television sector is predicted to be 10.6% in 2022
11	The Dynamics of OTT Platforms for Modern Indian Theatre	Survey data	Surveyed about changes in content production	Rapid development of OTT platforms poses a severe threat to traditional forms of entertainment.
12	Unstoppable Over-The-Top - An Overview of OTT Platform and Regulation in India	Survey data	Surveying data to understand the impact of Ott.	Legislation is required to address gaps in regulating content on OTT services.

## TOOLS & IDE USED

We carefully used a mix of strong tools and technologies to accomplish our project in order to increase productivity and efficiency. The two main IDEs that were used were Google Colab and Jupyter Notebook. These programmes are well-known for their interactive and collaborative capabilities, which allowed our team to work together seamlessly. We used well-known libraries like Numpy and Pandas to enable sophisticated data manipulation and analysis. These libraries were crucial in managing intricate datasets and carrying out necessary data operations. Moreover,

we used state-of-the-art visualization tools like Seaborn, Matplotlib, and Plotly to visualize our results and insights. These libraries allowed us to successfully explain our findings to stakeholders that were both technical and non-technical, in addition to offering a variety of visually appealing representation possibilities. Our project's success was greatly enhanced by the clever integration of various tools and software components, which streamlined the development, analysis, and communication processes.

## **METHODOLOGY**

1. **Data Preprocessing & Cleaning:** The kaggle streaming service's database has around sixteen thousand entries, making it rather huge. Here, doing EDA and obtaining insightful information is the goal. It might potentially be useful in predicting a film's rating. The results were not significantly affected after data cleaning as we removed some rows with missing values. Then, we have removed the Age and Rotten Tomatoes attributes because they appear to have a lot of missing i.e. NaN values. The Age feature later was revisited because there was still some helpful information out there. The imdb feature can handle the prediction (with just 3% of missing information), hence the 'Rotten Tomatoes' column was eliminated. A few rows with missing data, such as "directors," "titles," and so on, are eliminated; characteristics with multiple inputs, such as "Directors," "Genres," "Nation," and "Language," are handled. We create further columns to further split out these elements. Since there will be many more columns, separated dataframe fields were made for "Directors," "Genres," "Country," and "Language." Depending on the study, these data frames were joined.
2. **Exploratory data analysis (EDA):** is a procedure that involves conducting a preliminary analysis of the data to identify any abnormalities and shape it so that it may be used to get some insights into solving our goal. The first step in the pre-processing of the database of this study was to visualize the raw data using descriptive statistics tables, skewness, and other descriptive terms like mean, max,

and percentile values. It also involves preparing textual data for grouping purposes from user's text review and identifying and eliminating missing values. Then in cleaning of the textual data we have replaced missing values using different methods. Using imputation 'Empty string' was used to fill up the missing data in the director, cast, and country characteristics. There is a low percentage of nan values in the date\_added and rating columns; removing these values won't have a significant impact on the model's construction. Therefore, we only remove the nan value that is present in the date\_added and rating columns. We filled in the blanks with an empty string and have chosen to remove features that have fewer than 5% missing values outright. Additionally, the Capping method is used in the outlier removal procedure to eliminate outliers from the data where Q1, Q3 stand for each attribute's first and third quartiles as visualized(Fig.1).

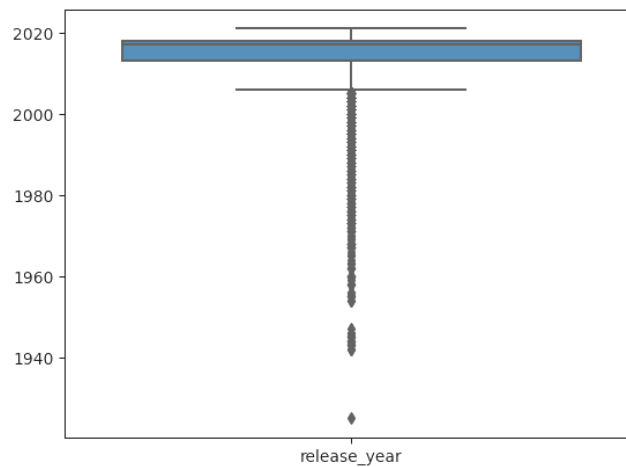


Fig.1 No Outlier Observed in the above boxplot

Prior to merging the metadata and expanding our analysis, a number of factors were taken into consideration, including which OTT platform has the most films, how the score distribution is on OTT platforms, how old the films are on OTT platforms, and how long the films are on OTT platforms.

3. Analysis: The presentation of the brand color palettes(Fig.2) for Disney, Netflix, and Amazon gives the study a visual element. These eye-catching color choices not

only help consumers recognise the brand, but they also give each streaming platform its own distinct personality and aesthetic. The distribution of content kinds (TV shows and films) for each platform is shown using bar graphs. Based on the preferences of the target audience, this study offers a thorough picture of the content environment and enables strategic decisions about content generation and purchase(Fig.4).

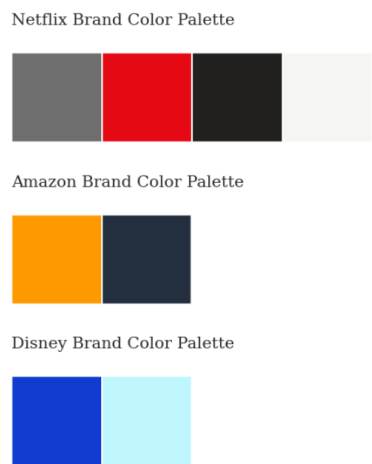


Fig.2 Brand color palettes for Disney, Netflix, and Amazon

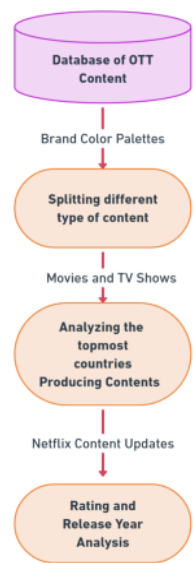


Fig.3 Block Diagram



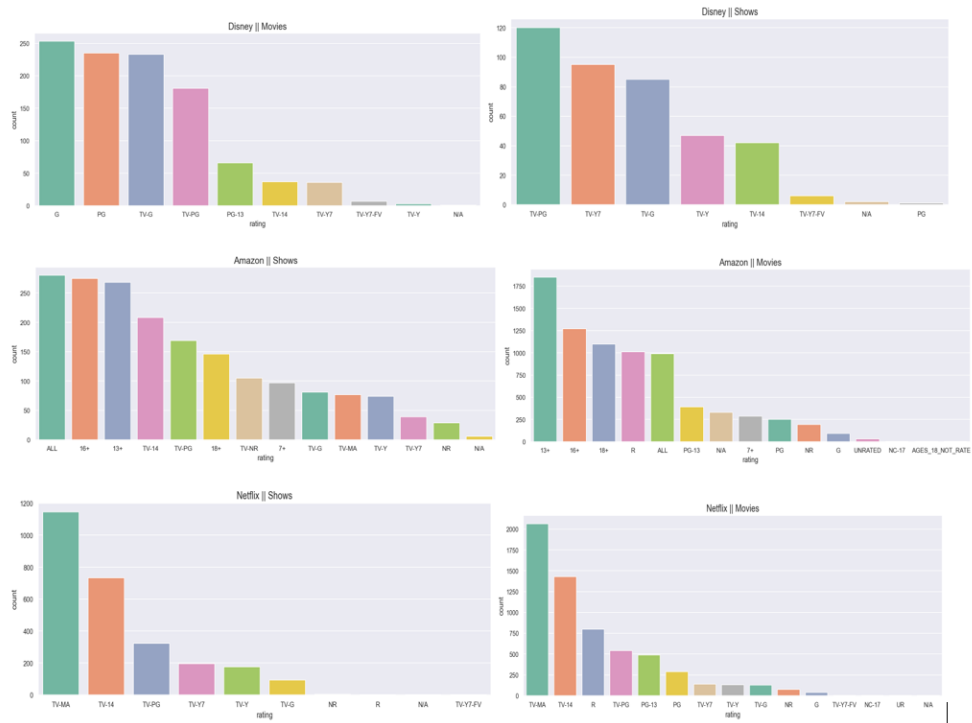


Fig.4 Distribution of content kinds

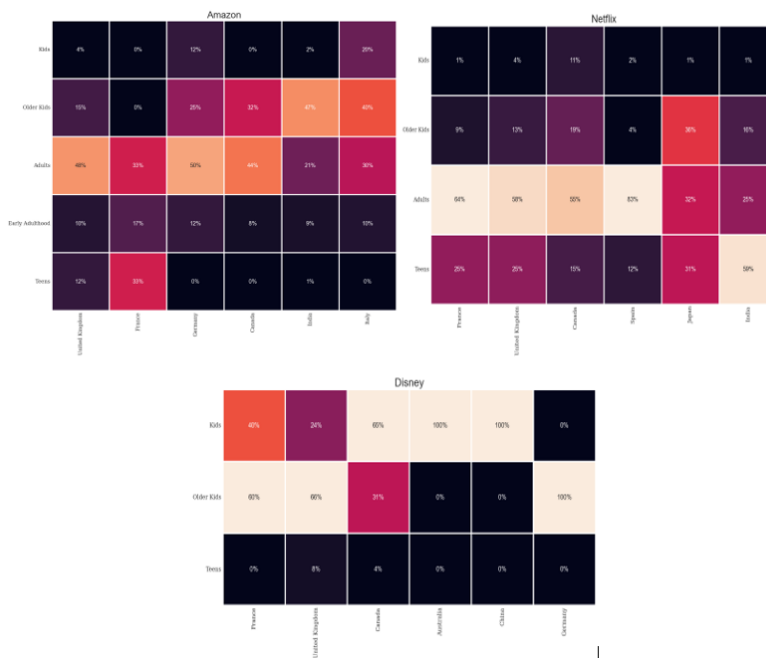


Fig.5 Heatmaps of Amazon prime, Hotstar and Netflix indicating age groups associated with particular OTT

Heatmaps indicate how frequently Netflix changes its material over time, with differentiating between TV series and films. Streaming systems may enhance user retention by maintaining a steady and captivating flow of new material by spotting patterns in content delivery and optimizing its schedule(Fig.5).

Bar plots provide insights into the diversity of content by displaying the distribution of content ratings for each platform. This data is essential for figuring out the audience's interests and creating future material that will appeal to a wide range of tastes. Bar charts illustrate the distribution of content release years, facilitating the identification of content creation patterns. Understanding how patterns in content creation change over time enables platforms to predict and adjust to evolving audience tastes and market conditions. The top 10 nations that produce the most films and TV series on each platform are displayed in bar plot style. This information helps streaming services explore new markets, develop foreign partnerships, and maintain a library of material that appeals to a worldwide audience. Based on ratings, heatmaps examine the target age groups for Disney, Netflix, and Amazon content. By comprehending the audience's demographics, platforms may improve their content strategy and make sure that it appeals to the target age groups and increases user happiness.

## **EXPERIMENTAL RESULTS**

While Amazon and Disney plus offer a comparatively bigger share of films, Netflix has a fair selection of both TV series and films. This knowledge directs content acquisition tactics, assisting platforms in keeping a varied and engaging content collection. Although Netflix regularly refreshes its programming, there are noticeable surges in some months that correspond to deliberate release schedules. By strategically and regularly updating its content, Netflix stands out in a crowded industry by increasing user engagement and loyalty. With material that spans various release years and a range of ratings, each platform serves a diversified

audience. Platforms are better able to reach a wider audience and compete in the market because of this variety.

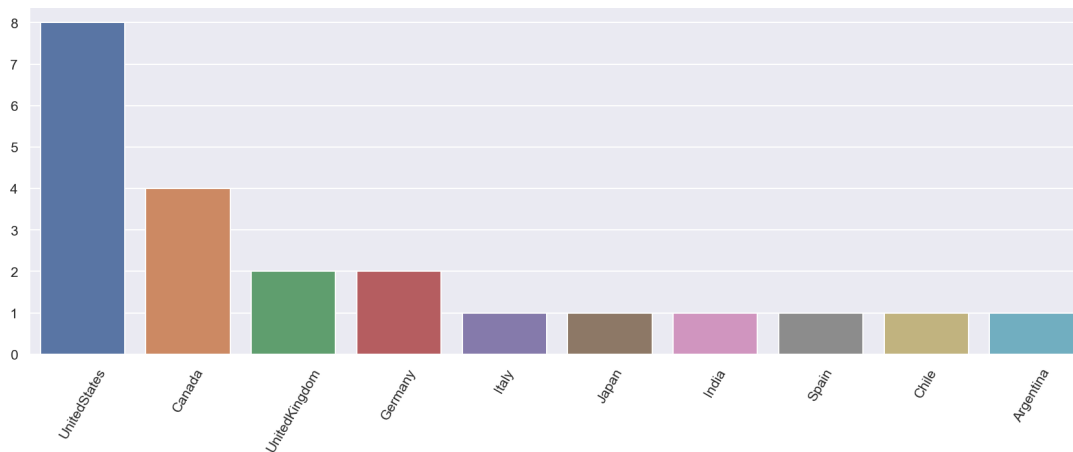


Fig.6 Bar Graph Showcasing the topmost 10 countries in content production

The USA, Canada, and the UK(Fig.6) stand out as major producers of content for all platforms. With the use of this information, platforms may concentrate on markets with strong content demand and cultural significance as part of their global expansion initiatives. Platforms target different age groups, and depending on ratings, they have different preferences for different types of material. Content that is age-appropriate increases viewer pleasure and loyalty, which helps streaming platforms survive in the competitive long run in the entertainment sector.

## CONCLUSION

The findings show us the practical insights for the growth and profitability of streaming platforms in the cutthroat entertainment sector, in addition to shedding light on the present content landscapes. Businesses may improve user happiness, draw in more customers, and hold a dominant position in the ever-changing entertainment industry by using these data.

## REFERENCES

- [1] S .Ghalawat et al., “Factors Influencing Consumer’s Choice of Streaming Over the Top (OTT) Platforms,” *Indian Journal of Extension Education*, Jan. 01, 2021: <https://doi.org/10.48165/ijee.2021.57323>
- [2] R. S. Dixit, S. L. Choudhary, and N. Govil, “Analyzing Consumer Preferences and Patterns for Ott Platforms: A Study,” Jan. 01, 2023. Available: <https://doi.org/10.2139/ssrn.4424186>
- [3] D. Patil, “Movies on Ott Analysis Using Multiple Regression and Random Forest in R,” [www.academia.edu](http://www.academia.edu), Jan. 01, 2021. [https://www.academia.edu/88475979/Movies\\_on\\_Ott\\_Analysis\\_Using\\_Multiple\\_Regression\\_and\\_Random\\_Forest\\_in\\_R](https://www.academia.edu/88475979/Movies_on_Ott_Analysis_Using_Multiple_Regression_and_Random_Forest_in_R)
- [4] K. Singharia, “Consumption of OTT Media Streaming in COVID-19 Lockdown: Insights from PLS Analysis,” [www.academia.edu](http://www.academia.edu), Nov. 16, 2022. [https://www.academia.edu/65308288/Consumption\\_of\\_OTT\\_Media\\_Streaming\\_in\\_COVID\\_19\\_Lockdown\\_Insights\\_from\\_PLS\\_Analysis](https://www.academia.edu/65308288/Consumption_of_OTT_Media_Streaming_in_COVID_19_Lockdown_Insights_from_PLS_Analysis)
- [5] R. G. Dastidar, “Pre and Post COVID-19 Sentiment Analysis of Consumers for OTT Platforms,” *Psychology*, Jan. 05, 2021.. Available: <https://doi.org/10.17762/pae.v57i9.2704>
- [6] P. Zhao, X. Gao, C. Xu, and L. Chen, “M5: Multi-Modal Multi-Interest Multi-Scenario Matching for Over-the-Top Recommendation,” Aug. 04, 2023. <https://doi.org/10.1145/3580305.3599863>
- [7] Kumar, S., Meena, R. P., & Scholar, R. (2023, August 11). *The Rise of OTT Platform: Changing Consumer Preferences*. ResearchGate. <https://doi.org/10.17605/OSF.IO/AN69G>
- [8] S. Bhoite, C. H. Patil, S. Pal3, and V. Magar4, “Polarity in sentiments using NLP: ML-based recommendation system.”
- [9] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, “A Survey on Multimedia Services QoE Assessment and Machine Learning-Based Prediction,” *IEEE Access*, vol. 10, pp. 19507–19538, 2022, doi: 10.1109/ACCESS.2022.3149592.
- [10] P. Zhao, X. Gao, C. Xu, and L. Chen, “M5: Multi-Modal Multi-Interest Multi-Scenario Matching for Over-the-Top Recommendation,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2023, pp. 5650–5659. doi: 10.1145/3580305.3599863.

- [11] A. Kurre and A. Rao, "Comparative Analysis: Machine Learning Usage Across Recommender Systems of OTT Platforms", doi: 10.58213/vidhyayana.v8i5.689.
- [12] N. Khare and A. Jhapate, "Review on Collaborative Filtering Machine Learning Approach for Recommendation Systems," 2022 [www.rjetm.in](http://www.rjetm.in)
- [13] Machine Learning Algorithms to Propose a Technique for Choosing the Efficient OTT Platform. (2023, June 7). International Research Journal of Modernization in Engineering Technology and Science. <https://doi.org/10.56726/irjmets41702>
- [14] E. Sundaravel and N. Elangovan, "Emergence and future of ZQZOver-the-top (OTT) video services in India: an analytical research," International journal of business, management and social research, Jan. 01, 2020 <https://doi.org/10.18801/ijbmsr.080220.50>