



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: <https://doi.org/10.22214/ijraset.2023.55802>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Netflix Movies and Tv Shows Data Analysis

Syed Reshma Banu¹, B. Sravani²

CSE, K L University, INDIA

Abstract: Netflix stands as the leading on-demand streaming platform in current times, offering its services across a staggering 190 countries and featuring an extensive library of both movies and TV shows. In our research, we carried out an initial exploratory analysis utilizing data sourced from Flexible, a search engine dedicated to cataloging the content accessible on the Netflix platform. When Netflix made its debut in April, online video streaming was still in its early stages. Fast forward eighteen years, and Netflix has evolved into the world's foremost global Internet television platform. We are diving into Netflix datasets obtained from Kaggle to gain a deeper understanding of the release schedules for movies and TV shows on the platform. Our goals encompass determining the time it takes for content to become accessible, scrutinizing the frequency of releases in specific time intervals, examining releases over the last decade, and identifying the top 10 genres that resonate most with Netflix viewers. The preliminary analysis has revealed fascinating observations regarding the current content trends on Netflix. While the current state of the recommendation system has its limitations, there is potential for improvement when incorporating additional features and variables.



I. INTRODUCTION

When we look at all the movies and TV shows available on Netflix, it gives us a peek into the ever-changing world of online streaming entertainment. Netflix is a pioneer in this field and has changed the way we watch movies and shows. Their collection is huge and includes many different types of movies and shows released at different times. The data we have from Netflix provides a special chance to discover important information about what kind of content Netflix offers and what viewers like. In this analysis, we want to dig into this data, looking at things like when they release content, how much they have, what's been popular over the years, and what types of shows and movies people enjoy the most on Netflix. We're going to use advanced data techniques to understand the story behind Netflix's content, its history, and maybe even where it's going in the future.

Netflix has continued to strengthen its position as a leading player in the world of streaming content, establishing an impressive foothold in the digital movie streaming landscape. Specifically, an overwhelming majority of digital movie streams, precisely six out of every ten, can be attributed to Netflix. This remarkable achievement was evident as of January 2012 when Netflix proudly reported a substantial subscriber base, boasting over 20 million loyal customers spread across 45 different countries worldwide. What's even more astonishing is the sheer scale of content consumption facilitated by the platform, with these subscribers collectively streaming an astounding 2 billion hours' worth of TV shows and movies. This level of engagement highlights the platform's popularity and its pivotal role in shaping the way people consume entertainment. On average, each user on the Netflix platform was consuming a considerable amount of data, surpassing a gigabyte per day. Such data-intensive viewing habits underline the platform's influence and significance in the modern media landscape, as well as its ability to seamlessly deliver content to a global audience.

From our initial examination of the dataset, it becomes evident that the recorded data spans from the earliest available content to the latest update in 2021. Within this dataset, we have access to a wealth of information encompassing details about directors, actors/actresses, movies, TV shows, ratings, content duration, and more. This extensive dataset equips us with the tools to explore various aspects of Netflix's operations.

We can use this data to discern how long it typically takes for Netflix to make content available on its platform, identify the most popular genres among viewers, determine which actors and actresses enjoy the most prominence, and analyze how genre popularity has influenced the choice of movies or TV shows in which an actor or actress appears over time. By employing the analytical capabilities of the R programming language, we aim to unravel the evolving trends within the Netflix platform, providing valuable insights into its content strategy and audience preferences.

In recent times, there has been a substantial upsurge in the number of TV shows accessible on the Netflix platform, with this expansion nearly tripling the available content. This notable growth has captured our interest and has motivated us to conduct a deeper exploration of the dataset, recognizing the potential wealth of additional insights that can be extracted from this rich source of data. Simultaneously, our ongoing efforts have culminated in the development of a sophisticated Netflix content recommendation system. This recommendation system relies on advanced similarity analysis techniques, specifically leveraging the power of Term Frequency-Inverse Document Frequency and Cosine Similarity algorithms. Initially, our analytical focus was centered on the textual components, namely the title and description of each piece of content. However, our future aspirations entail an ambitious plan to further enhance the recommendation system. We aim to broaden its scope by incorporating other pivotal determinants, with a particular emphasis on capturing and utilizing demographic information related to the platform's users. By doing so, we aim to create a more personalized and ultimately more effective content recommendation system that caters to the unique preferences and viewing habits of Netflix's diverse audience.

II. RELATED WORK

In the world of analyzing Netflix movies and TV shows, the main goal is to quickly provide users with helpful information using smart data analysis methods. In research, many efforts have tried to create effective analysis systems, often using a method called Cosine Similarity to compare things and spot trends. For example, there's a Netflix movie and TV show analysis system that you can use on websites. This system uses a simple data model or looks at images to help you find movies and TV shows on streaming platforms. Also, there's another analysis system that makes finding things easier by using Machine Learning and Cosine Similarity. It considers things like how much people like something, what type of show it is, and what it's about to give you meaningful information, helping you make better choices when picking what to watch.

III. PROPOSED METHODOLOGY

A. Data Collection

To source the data for your analysis, you can consider various options. One approach is to gather relevant data related to movies and TV shows, including viewers' viewing histories, ratings, and reviews, as well as metadata such as information about the cast and crew. Some publicly available datasets, like those released by researchers and organizations, provide comprehensive information related to Netflix content, including user interactions and ratings. Notable examples include the Netflix prize dataset and datasets resulting from academic research. Alternatively, if Netflix offers an Application Programming Interface (API) for developers, you may have the opportunity to access real-time or historical data directly from the platform. This can be particularly valuable for obtaining up-to-date information. Another approach is web scraping, where you extract data directly from the Netflix website. This can include details such as titles, descriptions, cast and crew information, user reviews, and more. However, it's essential to note that web scraping requires coding skills and adherence to ethical scraping practices and the website's terms of service.

B. Loading The Data Set

Loading the dataset for Netflix movie and TV shows data analysis is a critical initial step in the analytical process. This dataset typically contains a wealth of information, including details about the content available on Netflix, user interactions, ratings, and more. To begin, data analysts or researchers acquire the dataset from relevant sources, which may include publicly available datasets, Netflix's API if accessible, or web scraping from the Netflix website.

Once obtained, the dataset is imported into the chosen data analysis environment, which could be a programming language like Python or a data analysis tool like R. Analysts then perform data validation, ensuring that the dataset is complete and well-structured. This involves checking for missing values, handling duplicates, and addressing any data inconsistencies or anomalies.

The loaded dataset's structure is examined, including the number of rows and columns, data types, and initial summary statistics. This preliminary exploration helps analysts understand the dataset's size and composition.

After loading and basic validation, the dataset is ready for exploratory data analysis (EDA), where analysts dive deeper into its contents to uncover patterns, trends, and insights. EDA involves various techniques such as data visualization, summary statistics, and data transformation to prepare the data for more advanced analysis, modeling, and decision-making in the context of Netflix movie and TV shows data analysis.

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---------|---------|-------------------------------------|---------------------|-------------------------------|----------------|----------------|--------------|--------|-----------|-------------------|---|
| s1 | Movie | Dick Johnson | Kirsten Johnson | | United States | September 2020 | 2020 | PG-13 | 90 min | Documentary | As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in |
| s2 | TV Show | Blood & Water | Ama Qam | South Africa | September 2021 | 2021 | TV-MA | | 2 Seasons | International | After crossing paths at a party, a Cape Town teen sets out to prove whether a private-s |
| s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajila, Tracy | September 2021 | 2021 | TV-MA | | 1 Season | Crime TV | To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team |
| s4 | TV Show | Jailbirds | New Orleans | | September 2021 | 2021 | TV-MA | | 1 Season | Documentary | Feuds, flirtations and toilet talk go down among the incarcerated women at the Orleans |
| s5 | TV Show | Kota Factory | Mayur Mehta | India | September 2021 | 2021 | TV-MA | | 2 Seasons | International | In a city of coaching centers known to train India's finest collegiate minds, an earn |
| s6 | TV Show | Midnight Mass | Mike Flanigan | Kate Siegel, Zach Gilford | September 2021 | 2021 | TV-MA | | 1 Season | TV Drama | The arrival of a charismatic young priest brings glorious miracles, ominous mysteries ar |
| s7 | Movie | My Little Pony: A New Generation | Robert Cullen | Vanessa Hudgens, Kiara Glasco | September 2021 | 2021 | PG | | 91 min | Children & Family | Equestria's divided. But a bright-eyed hero believes Earth Ponies, Pegasi and Unicorns s |
| s8 | Movie | Sankofa | Haile Geri | Kofi Ghani | United States | September 1993 | 1993 | TV-MA | 125 min | Drama | On a photo shoot in Ghana, an American model slips back in time, becomes enslaved o |
| s9 | TV Show | The Great British Bake Off | Andy Dewhurst | Mel Giedroyc, Noel Fielding | United Kingdom | September 2021 | 2021 | TV-14 | 9 Seasons | British TV | A talented batch of amateur bakers face off in a 10-week competition, whipping up the |
| s10 | Movie | The Starling | Theodore | Melissa McCarthy | United States | September 2021 | 2021 | PG-13 | 104 min | Comedies | A woman adjusting to life after a loss contends with a feisty bird that's taken over her g |
| s11 | TV Show | Vendetta: Truth, Lies and The Mafia | | | September 2021 | 2021 | TV-MA | | 1 Season | Crime TV | Sicily boasts a bold "Anti-Mafia" coalition. But what happens when those trying to bring |
| s12 | TV Show | Bangkok | Bongkollawat | Kanarot | September 2021 | 2021 | TV-MA | | 1 Season | Crime TV | Struggling to earn a living in Bangkok, a man joins an emergency rescue service and rea |
| s13 | Movie | Je Suis Karim | Christian Schwochow | Luna Wedderburn | Germany | September 2021 | 2021 | TV-MA | 127 min | Drama | After most of her family is murdered in a terrorist bombing, a young woman is unknow |
| s14 | Movie | Confession | Bruno Garcia | Klara Castanho, Lucca Pimenta | September 2021 | 2021 | TV-PG | | 91 min | Children & Family | When the clever but socially-awkward Tetã joins a new school, she'll do anything to fit |
| s15 | TV Show | Crime Stories: India Detectives | | | September 2021 | 2021 | TV-MA | | 1 Season | British TV | Cameras following Bengaluru police on the job offer a rare glimpse into the complex ar |
| s16 | TV Show | Dear White People | Logan Brown | United States | September 2021 | 2021 | TV-MA | | 4 Seasons | TV Comedy | Students of color navigate the daily slights and slippery politics of life at an Ivy League c |
| s17 | Movie | Europe's Dark | Pedro de Alcntara | Garca, Pablo | September 2020 | 2020 | TV-MA | | 67 min | Documentary | Declassified documents reveal the post-WWII life of Otto Skorzeny, a close Hitler ally w |
| s18 | TV Show | Falsa identidad | Luis Ernesto | Mexico | September 2020 | 2020 | TV-MA | | 2 Seasons | Crime TV | Strangers Diego and Isabel flee their home in Mexico and pretend to be a married coup |
| s19 | Movie | Intrusion | Adam Salkov | Freida Pinto, Logan Lerman | September 2021 | 2021 | TV-14 | | 94 min | Thrillers | After a deadly home invasion at a couple's new dream house, the traumatized wife |

C. Data Cleaning

Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.

Standardize data formats and ensure data quality for analysis.

Before you start analyzing data, it's important to get it in good shape. This means dealing with problems like missing information, strange values, and mistakes in the data. You also need to make sure the data is in a format that makes sense for analysis.

```
print("\nColumns with missing value:")
```

```
print(netflix_df.isnull().any())
```

Columns with missing value:

```
show_id      False
type         False
title        False
director     True
cast         True
country      True
date_added   True
release_year False
rating       True
duration     False
listed_in    False
description  False
dtype: bool
```

One common problem is when you have the same data repeated more than once in your dataset. This can mess up your analysis. To fix this:

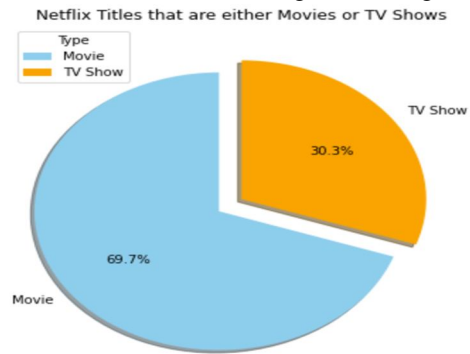
Find and get rid of the duplicate rows, either by looking at specific parts of the data or checking the whole thing.

But be careful when you do this because sometimes, there might be duplicates that are actually important.

This cleaning and getting the data ready part is really important to make sure your analysis is accurate and useful.

D. Exploratory Data Analysis (EDA)

Performing Exploratory Data Analysis (EDA) is essential for obtaining initial insights from the dataset.

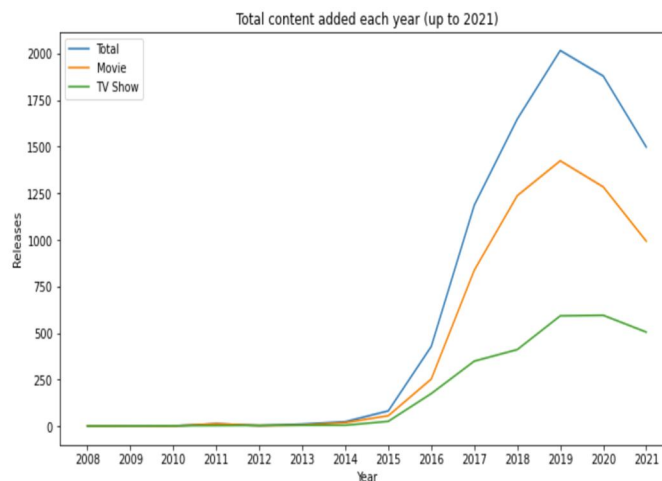
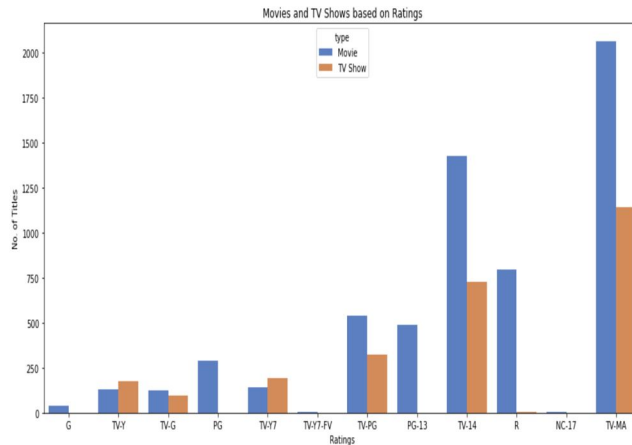


IV. DATASET OVERVIEW AND SUMMARY

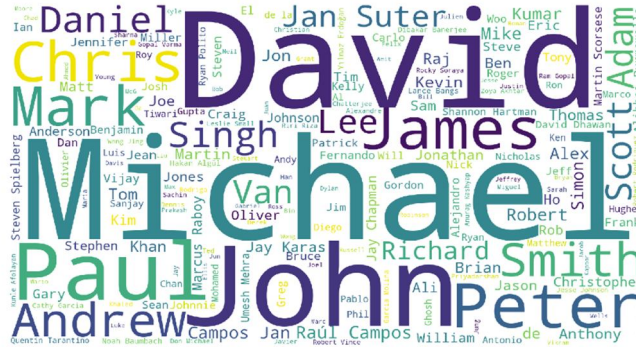
To begin, it's crucial to acquire a comprehensive overview of the dataset. This includes understanding its size, structure, and basic statistics. You'll want to know how many rows and columns it contains, the data types of each column, and key summary statistics for numerical variables, such as average, middle value, and how spread out the data is.

A. Data Visualization

To make underlying patterns and trends more visible, leverage a range of data visualization techniques. These include various types of charts and plots, like bar charts, line graphs, scatter plots, pie charts, histograms, and heatmaps. These visual representations help in gaining a deeper understanding of the data.



B. Word Cloud Representation



V. DISCUSSION AND CONCLUSION

At present, Netflix stands as the preeminent on-demand streaming platform, delivering a vast array of content to subscribers across 190 countries globally. From its inception, movies have held sway as the primary content in the realm of online programming. Remarkably, in 2020, there was a notable shift as TV shows took center stage, a trend that continued into the early months of 2021. In our research, we carried out an initial exploratory analysis of Netflix data. This analysis brought to light several key findings, such as the prominent availability of Netflix content in countries like the USA, India, and the United Kingdom. It also revealed the distribution of programs by category, with movies making up 69% and TV shows comprising 31%. Additionally, we conducted a semantic analysis of the descriptions of the works available on the platform.

We also looked at where Netflix content is popular, with countries like the USA, India, and the United Kingdom being the top places. Additionally, we analyzed the words used to describe the shows and movies, which helped us understand what they are about. As Netflix keeps influencing the entertainment world, our research shows that being flexible and using data to make decisions are essential in the fast-changing world of online streaming. These insights can be helpful for both the people who make the content and the people who watch it, especially in a time when digital media is so important

However, some may question the reliability of the results from this system due to the limited amount of data in the corpus we used. Nonetheless, it can serve as a good starting point for exploring other factors that could make recommendations even better, such as the duration of the shows, the ratings given on Netflix, the main actors in the program, and more. Additionally, we believe it would be valuable to include information about the subscribers' demographics, which Netflix currently doesn't consider when making recommendations. Despite its simplicity and the small amount of data needed for this recommendation system, it has the advantage of being easy to set up and use.

REFERENCES

- [1] <https://www.scitepress.org/Papers/2021/107275/107275.pdf>
- [2] Netflix Is Getting Its Own Cable Channel | Business | WIRED. Wired.com. Retrieved May 8, 2014, from <http://www.wired.com/2014/04/netflix-cable/>
- [3] How does Netflix recommend movies? Retrieved May 3, 2014, from http://scenic.princeton.edu/network20q/wiki/index.php?title=Q4:_How_does_Netflix_recommend_movies%3F
- [4] Taube, A. (2014, April 16). Proof That Netflix Is Destroying Cable TV. Business Insider. Retrieved May 3, 2014, from <http://www.businessinsider.com/netflix-eating-cable-subscriptions-experian-2014-4#!HJHFt>
- [5] Ozer, J. (2011, February 26). What is Streaming? – Streaming Media Magazine. Streaming Media Magazine. Retrieved May 3, 2014, from <http://www.streamingmedia.com/Articles/ReadArticle.aspx?ArticleID=74052>
- [6] Zambelli, A. (2013, March 1). A history of media streaming and the future of connected TV. theguardian.com. Retrieved May 3, 2014, from <http://www.theguardian.com/media-network/media-network-blog/2013/mar/01/history-streaming-future-connected-tv>
- [7] Netflix Is Getting Its Own Cable Channel | Business | WIRED. Wired.com. Retrieved May 8, 2014, from <http://www.wired.com/2014/04/netflix-cable/>
- [8] How does Netflix recommend movies? Retrieved May 3, 2014, from http://scenic.princeton.edu/network20q/wiki/index.php?title=Q4:_How_does_Netflix_recommend_movies%3F
- [9] Taube, A. (2014, April 16). Proof That Netflix Is Destroying Cable TV. Business Insider. Retrieved May 3, 2014, from <http://www.businessinsider.com/netflix-eating-cable-subscriptions-experian-2014-4#!HJHFt>
- [10] Ozer, J. (2011, February 26). What is Streaming? – Streaming Media Magazine. Streaming Media Magazine. Retrieved May 3, 2014, from <http://www.streamingmedia.com/Articles/ReadArticle.aspx?ArticleID=74052>
- [11] Zambelli, A. (2013, March 1). A history of media streaming and the future of connected TV. theguardian.com. Retrieved May 3, 2014, from <http://www.theguardian.com/media-network/media-network-blog/2013/mar/01/history-streaming-future-connected-tv>
- [12] https://www.researchgate.net/publication/359747031_Data_Analysis_on_Netflix_datasets



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)