NETFLIX DATA ANALYSIS

DATA 230 – Data Visualization
GROUP 4

TABLE OF CONTENTS

Abstract

A thorough data set on Netflix original TV series and movies found in the dataset "Netflix Shows - Exploratory Analysis" on Kaggle. It contains information on more than 7,000 entries, including each title's title, kind, director, cast members, country of origin, release year, rating, duration, and category. By using statistical and data visualization approaches, the exploratory analysis's goal is to gain insights from the dataset. It analyzes trends in the addition of titles over time, identifies the top contributing nations and genres, and looks into correlations between various characteristics like rating and duration. It also explores the distribution of variables like type, rating, and country of origin. The research also offers a thorough overview of the most well-known titles under different categories. This dataset and analysis are useful for comprehending the Netflix content products and services, determining audience preferences, and showcasing the power of data visualization and statistics for deriving insightful conclusions from large, complicated datasets.

1. INTRODUCTION

Netflix is one of the most well-known participants in this market. The emergence of streaming platforms has completely changed the way we consume entertainment. Netflix has attracted fans with its wide range of content options, including a sizable collection of original TV series and films from around the globe. It is essential for content producers, marketers, and data analysts to comprehend the trends, patterns, and qualities of this content. A thorough collection of data on Netflix original series is offered via the dataset "Netflix Shows - Exploratory Analysis" on Kaggle, enabling thorough exploration and analysis.

Titles, types (TV show or movie), directors, cast members, countries of origin, release years, ratings, durations, and categories are just a few of the details included in the collection. With more than 7,000 items, it provides a comprehensive and varied snapshot of Netflix's content environment. For conducting exploratory analysis to learn more about Netflix's content strategy, geographical variances in material, genre preferences, and temporal patterns, this dataset is an invaluable resource.

With the help of statistical and data visualization approaches, we hope to do an exploratory analysis of the "Netflix Shows" dataset in this project. By doing so, we hope to be better able to understand the platform's content. We can spot major trends and features of Netflix's content portfolio by looking at the distribution of variables like kind, rating, and country of origin. We will look into the library's development through time, examining the addition of new titles and their distribution across various genres and categories. We will also look at how different factors relate to one another, such as the relationship between ratings and runtimes or the appeal of particular genres in various nations.

1.1 PROJECT BACKGROUND

With the introduction of streaming services, the entertainment business has undergone a substantial upheaval, and Netflix has established itself as a major player in this new digital environment. Netflix has a sizable international audience because of its extensive library of original TV series and films. For content producers, marketers, and analysts, it is crucial to comprehend the patterns, trends, and characteristics of Netflix's content. A comprehensive collection of data on Netflix original series is provided through the "Netflix Shows - Exploratory Analysis" dataset, which is accessible through Kaggle. This allows users to explore and analyze Netflix's content offering. You may learn a lot about content strategy, regional differences, and audience preferences by looking at parameters like title, type, director, cast, country of origin, release year, rating, runtime, and category.

1.2 PROBLEM DEFINITION

To better understand Netflix's content availability and the underlying patterns in the dataset, this project will do an exploratory analysis of the "Netflix Shows - Exploratory Analysis" dataset. The initiative aims to derive useful insights that might guide content producers, marketers, and analysts in their decision-making processes by using statistical analysis and data visualization tools. The project will involve analyzing the distribution of variables, looking at temporal trends, and determining how the variables in the dataset relate to one another. The end goal is to use data analysis to discover audience preferences, gain important insights into Netflix's content environment, and support data-driven decision-making in the entertainment sector.

1.3 OBJECTIVE

The primary objective of this project is to examine the Netflix dataset and use data visualization techniques to convey important insights about the performance of the platform. A thorough understanding of the service can be attained by looking at user demographics, content choices, and viewing patterns. The conveyance of difficult information will be sped up through the use of data visualization, making it simple for stakeholders to comprehend. The ultimate objective is to offer insightful, data-driven knowledge that may assist marketers, industry analysts, and content producers in making decisions.

1.4 Literature Review

In the paper "Exploratory and Sentiment Analysis of Netflix Data " by Karthik Babu Vadloori and Shriya Madhavi Sanghishetty is on these two aspects of the study. Although this paper's exact literature review is not readily available, it is reasonable to assume that it will examine pertinent research and methodologies for exploratory analysis and sentiment analysis in the context of Netflix data. The techniques used to assess the feelings conveyed in Netflix data include text mining, machine learning, and natural language processing, among others. Insights into market research, content selection algorithms, and customer feedback are among the goals of the authors' study, which will help us better understand Netflix users' attitudes and preferences.

Netflix has become a major supplier of entertainment material globally due to the fast expansion of streaming services and the growing significance of tailored content suggestions. With an emphasis on trends, techniques, and insights, this literature review examines the current state of research on Netflix recommendation systems. The poll opens with a description of Netflix as a business and a discussion of its enormous influence on the entertainment sector.

The usage of Big Data and recommendation systems in relation to Netflix's content selection and user preferences is then covered in depth. The survey also describes the data gathering and preparation techniques used, including the usage of the TF-IDF and Cosine Similarity algorithms to provide recommendations. An exploratory data study of Netflix's programming offerings, including country rankings, is provided in the findings section. The poll then explores the ramifications of the results, points up relevant works in the area, and offers suggestions for further research.

The paper "Netflix Big Data Analytics - The Emergence of Data Driven Recommendation" by Srivatsa Maddodi, & Krishna Prasad, K presents a case study that examines how Netflix uses big data analytics in its recommendation system. The study draws information from various sources, including the Netflix website, blogs, and academic articles. The objectives of the case study include understanding Netflix's history and evolution, analyzing their strategies in overcoming competition, examining the recommendation system, and exploring the application of big data analytics to enhance customer satisfaction. The paper discusses Netflix's early adoption of big data analytics, including their participation in the Netflix Prize competition to improve their recommendation algorithm. The methodology involves a thorough analysis of Netflix's business model, the types of recommendation systems used (content-based, collaborative, and hybrid), and the limitations faced. Overall, the study provides insights into how Netflix leverages big data analytics to personalize recommendations for its users.

2. Problem Statement

By analyzing the data and coming up with insights, Netflix may decide what kinds of series and movies to make and how to expand their company in different regions.

- What is the most producing genre?
- Country wise content on Netflix ?
- Average ratings on Netflix?
- Top directors in terms of films and TV shows?
- Does Netflix now place more of an emphasis on TV shows than on movies?
- Ideal month to release movies.

3. Motivation

Netflix's massive user base presents analysts and researchers with a unique challenge due to the sheer volume of data generated, which stood as a challenging task that motivated us to choose this topic.Using Data Visualization technology, by creating visual representations of the data, we can identify trends, patterns, and insights, which are often looked over.In order to understand Netflix's success, we can use data visualization to learn about user behavior and content performance.We hoped and were able to explore popular genres, audience retention rates, and the impact of original content through this project.

4. CRISP-DM METHODOLOGY

A widely used framework for data mining and analytics projects is the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining). It offers an organized method for guiding each of the processes required in deriving insightful information from data. The CRISP-DM methodology's individual parts are described in full below:

1.Data Understanding:

Finding the appropriate data sources, gathering the data, and performing exploratory data analysis (EDA) to evaluate its completeness, quality, and structure are all necessary steps in this process. The team wants to find any problems or restrictions with the data that might affect later phases of the project.

2. Business Understanding:

During this phase, stakeholders and subject-matter experts will work together to define the project's business objectives and goals. The team aims to have a thorough awareness of the constraints, requirements, and business environment. It is critical to coordinate data mining operations with the organization's unique goals and priorities. The project's subsequent phases are guided by clear commercial objectives, which also guarantee that the analysis is focused and substantive.

3. Data Preparation:

To make sure the data is suitable for analysis, data preparation entails cleaning, converting, and integrating the data. Managing missing values, addressing outliers, normalizing variables, and choosing pertinent subsets of data are a few examples of this. Additionally, derived variables must be developed, and data must be aggregated as necessary. In order to prepare the raw data for further modeling and analysis, it must be transformed into a well-structured dataset.

4. Modeling:

To create predictive or descriptive models, a variety of approaches and algorithms are applied to the prepared dataset during the modeling phase. Choosing the best modeling approaches depends on the goals of the project and the features of the dataset. Cross-validation and holdout validation are two strategies used to train and evaluate the models. The objective is to create models that successfully handle the company objectives and offer precise forecasts or insightful information.

5. Evaluation:

The evaluation phase is concerned with determining the efficiency and performance of the created models. This entails reviewing the findings, evaluating the model's precision and robustness, and contrasting it with the specified evaluation standards and corporate goals. The evaluation procedure aids in identifying any shortcomings or potential growth areas in the models, enabling modification and iteration as needed.

6. Deployment:

The insights, models, or analytical findings are included into the business procedure or system during the deployment phase. In order to achieve the desired business outcomes, this may entail operationalizing the findings, generating dashboards or reports for decision-makers, or integrating the models into production systems. Monitoring the performance of the deployed models and making the necessary corrections as new data becomes available are also part of the deployment process.

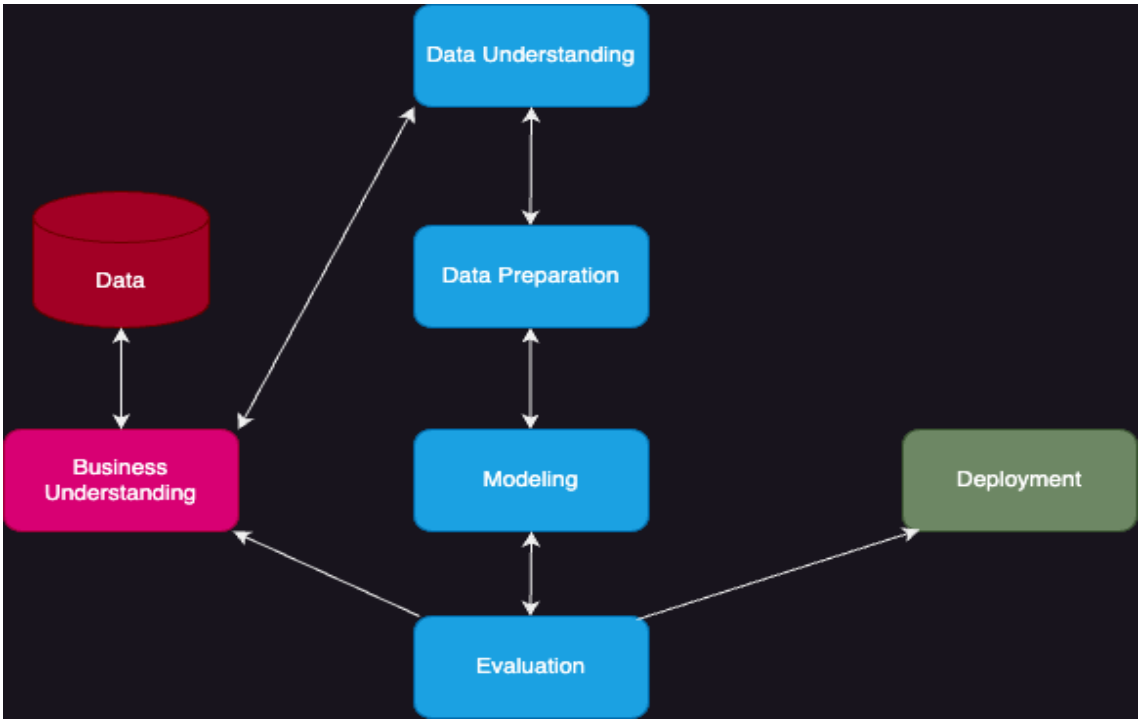


Figure 1: CRISP- DM Diagram

5. DATA FLOW DIAGRAM

Understanding the data and its sources comes before designing a data flow diagram. This entails determining the data types needed for the project, as well as its sources and formats. To ensure accuracy and relevance in the following processes, it is essential to have a thorough grasp of the data being used. Data cleansing and transformation come next after the data has been understood. This entails checking the data for any discrepancies, mistakes, or missing numbers before making any necessary corrections. It is possible to aggregate or de-aggregate data, convert data between different formats, or create new variables by performing mathematical operations to the data.

Following data cleansing and transformation, the objective of data enhancement is to add more information to the dataset. Data from several sources may be taken together, external data sets may be used, or derived variables may be produced based on calculations or business requirements. The goal of data enhancement is to raise the level of detail and quality of the data, increasing its value for analysis and decision-making. The project enters the data analytics phase after the data has been improved. In this step, numerous analytical methods and algorithms are applied to the data in order to draw conclusions, spot trends, and make predictions. The project team can discover important data using data analytics and come to conclusions that help the project achieve its goals.

Last but not least, data visualization is essential for conveying the results of data analytics. To effectively communicate the findings, this stage entails generating visual representations like charts, graphs, and dashboards. Stakeholders may more easily understand the main messages thanks to data visualization, which makes it easier to understand complicated linkages, trends, and patterns within the data. Decision-makers can also use it to rapidly find and understand the data that is pertinent to their needs. Data visualization increases the project's effect and facilitates data-driven decision-making by giving the data a visual representation. The data flow diagram, in its whole, offers an organized method for comprehending and managing data within a project workflow, facilitating effective data processing, analysis, and communication.

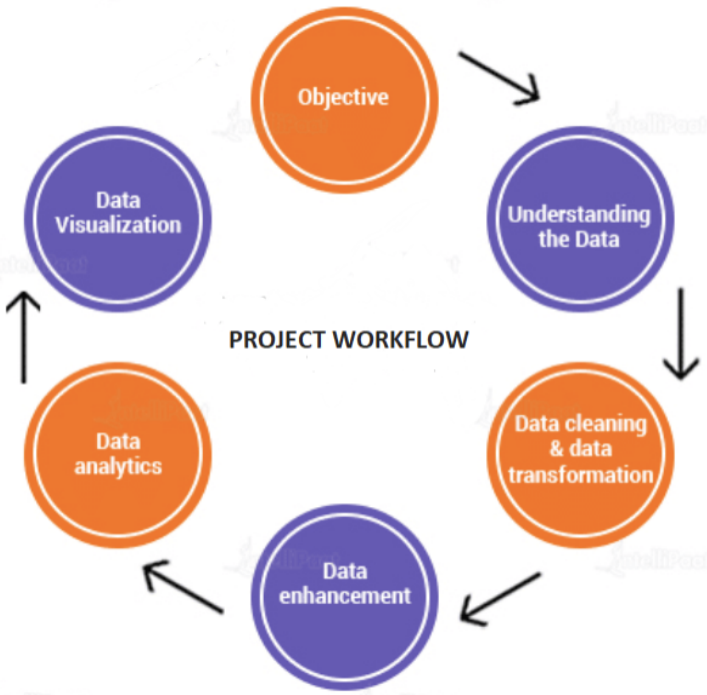


Figure 2: Data Flow Diagram

6. DATA COLLECTION

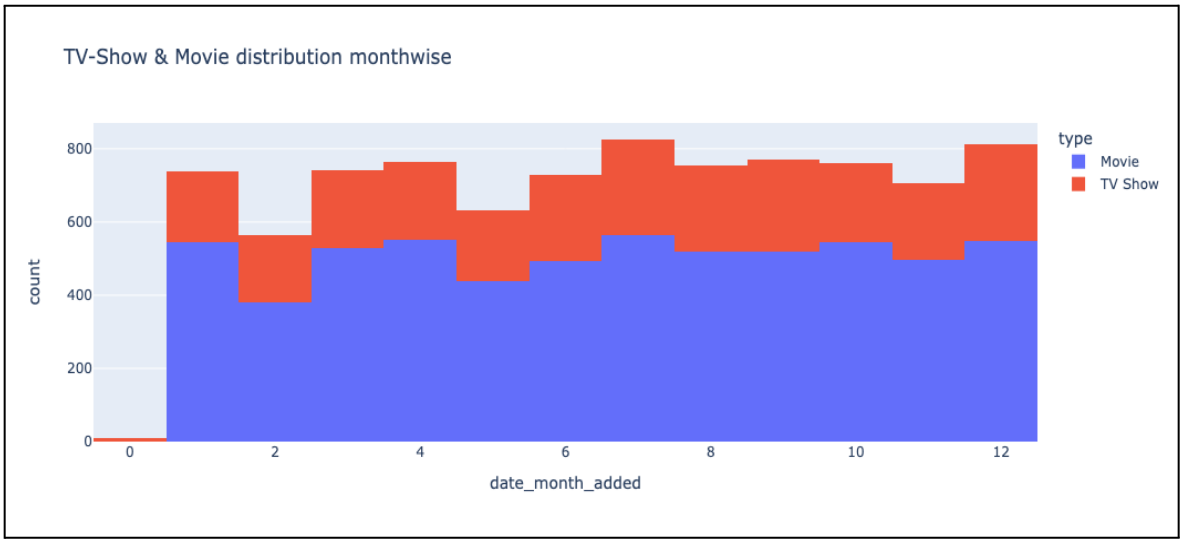
This research attempted to assure a thorough and objective study by collecting data from many sources, including Kaggle and web scraping using APIs. Due to the use of two datasets, thorough preprocessing and detailed visualizations were possible, leading to important discoveries that would not have been possible with just one data source. This method improves the validity and dependability of the results and helps us understand the Netflix platform more comprehensively.

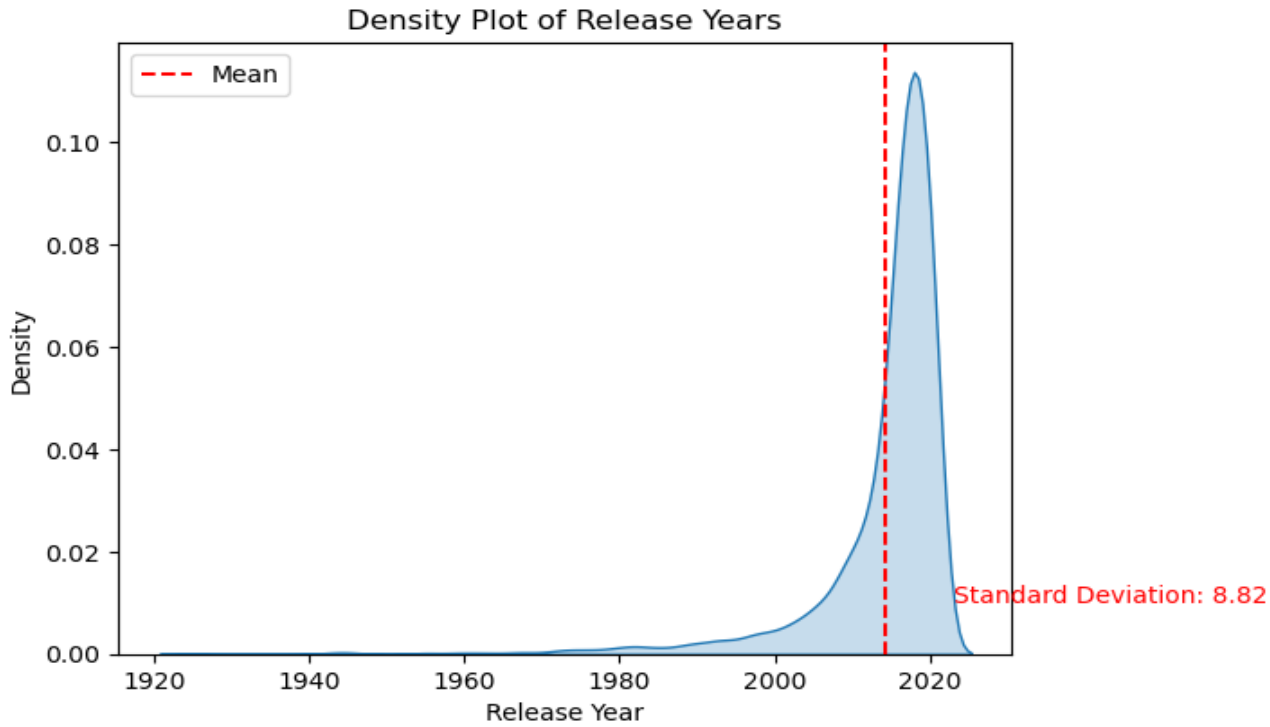
The initiative reduced the danger of depending just on one survey or dataset by mixing data from other sources, which decreased the possibility of bias and increased the robustness of the results reached. The research was further enhanced by the integration of many data sources since it gave a more comprehensive view of user preferences, habits, and geographical differences.

7. DATA PRE-PROCESSING

7.1 EDA

EDA Data Distribution : Left Skewed and Evenly Skewed





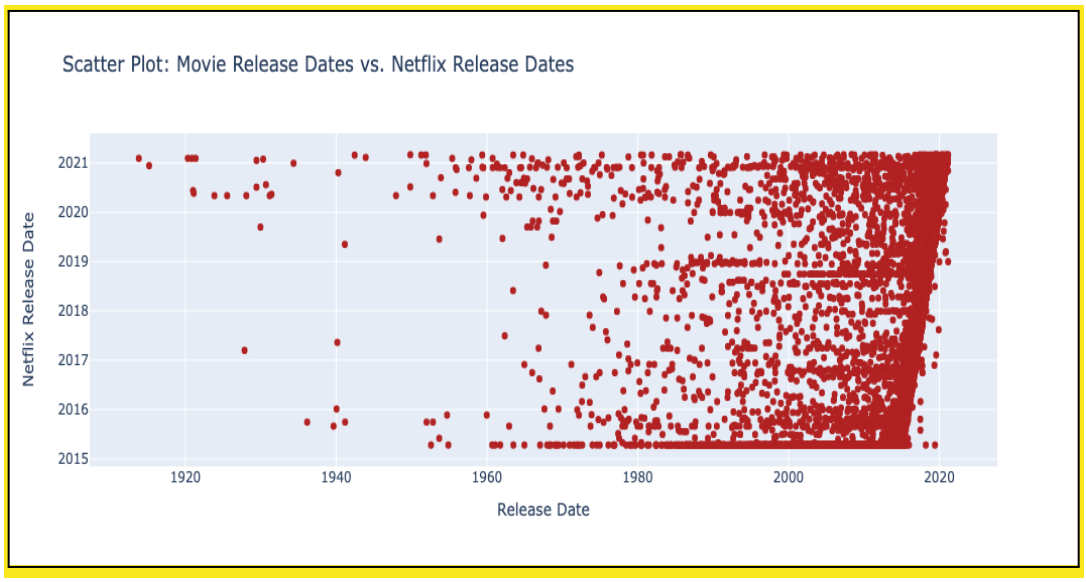
★ Our dataset is evenly distributed among all the months so, it won't cause biased decisions when insights are explored.

★ The Leftmost graph shows the left skewed distribution based on the year, Basically, it is due

to the fact that, netflix may not have the movies or TV-shows which are less popular from that time.

★ The Rightmost graph shows the even distribution of data TV-Show & Movie distribution monthwise.

EDA: Associations among release dates & correlation

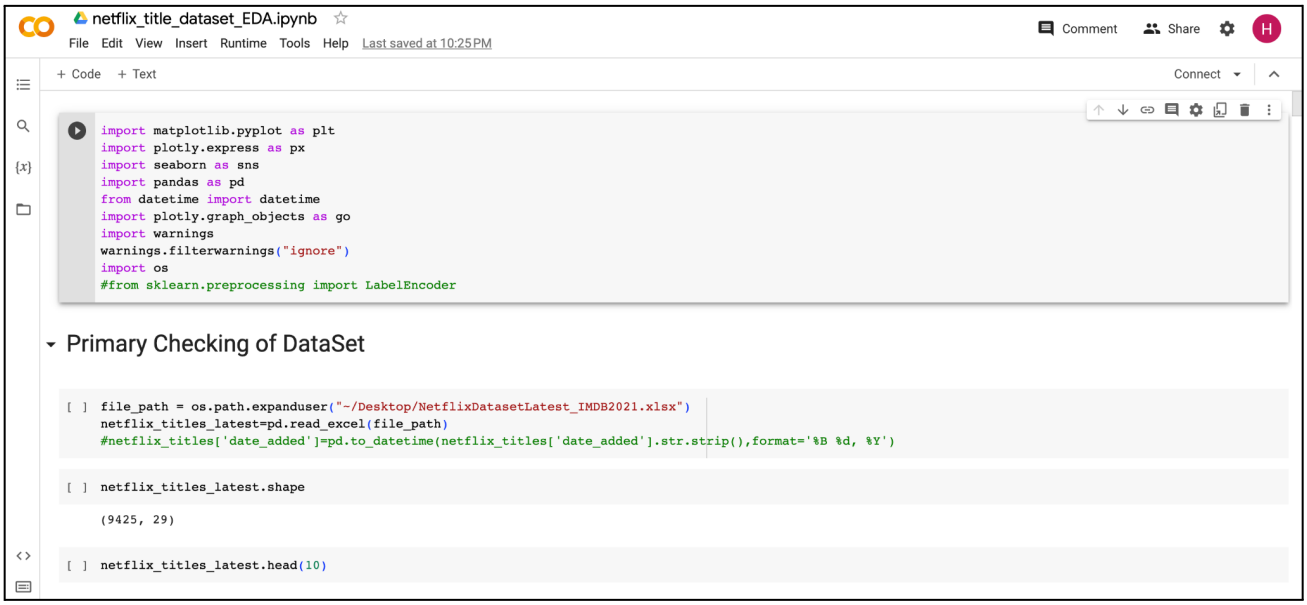


★ The graph provides the association between Movie release date vs netflix release date, this helps us to find the trend between the actual release date and netflix release date. These days, the movies are releasing after 90 days and some movies are releasing directly on Netflix and there are expected to be released within or next month.

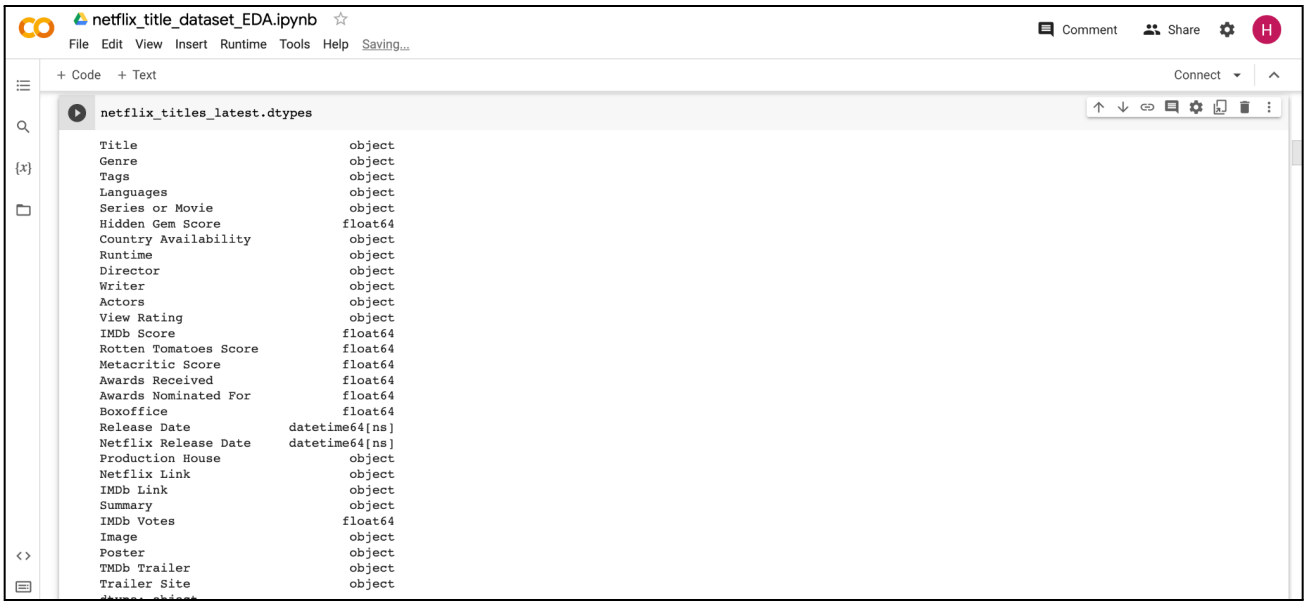
7.2 Data Cleaning

We have taken another dataset to explore from the flixgem website which contains Netflix TV Shows information like Title', 'Genre', 'Tags', 'Languages', 'Series or Movie', 'Hidden Gem Score', 'Country Availability', 'Runtime', 'Director','Writer', 'Actors', 'View Rating', 'IMDb Score','Rotten Tomatoes Score', 'Metacritic Score', 'Awards Received','Awards Nominated For', 'Boxoffice', 'Release Date','Netflix Release Date', 'Production House', 'Netflix Link', 'IMDb Link','Summary', 'IMDb Votes', 'Image', 'Poster', 'TMDb Trailer', 'Trailer Site', 'release years', 'Netflix

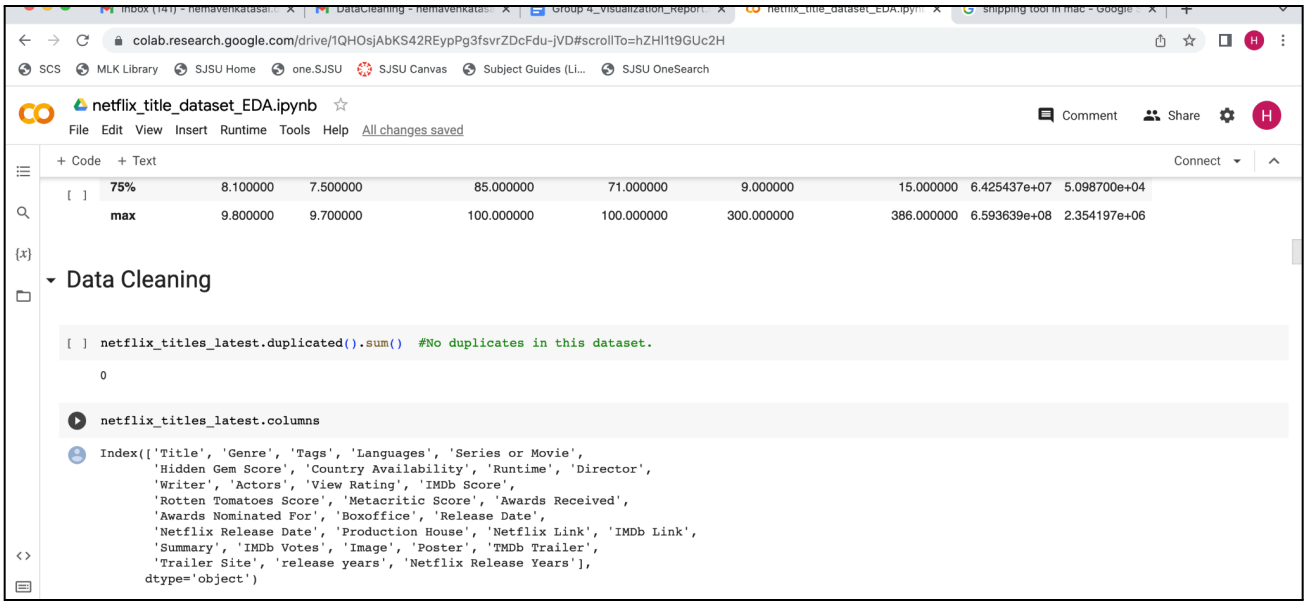
Release Years'.Initially, We imported required packages which are related to handling the data which is stored in our local system. The second dataset contains 29 columns and 9425 rows which are sufficient for performing the operations on the data to get the insights in various aspects.



Step i) First of all, We checked the data type of each column. Our columns are properly aligned with required data types. So, There is no specific requirement that the data type change appropriately.



Step 2: In the second level of cleaning, We checked for duplicate columns & rows but data doesn't contain any.

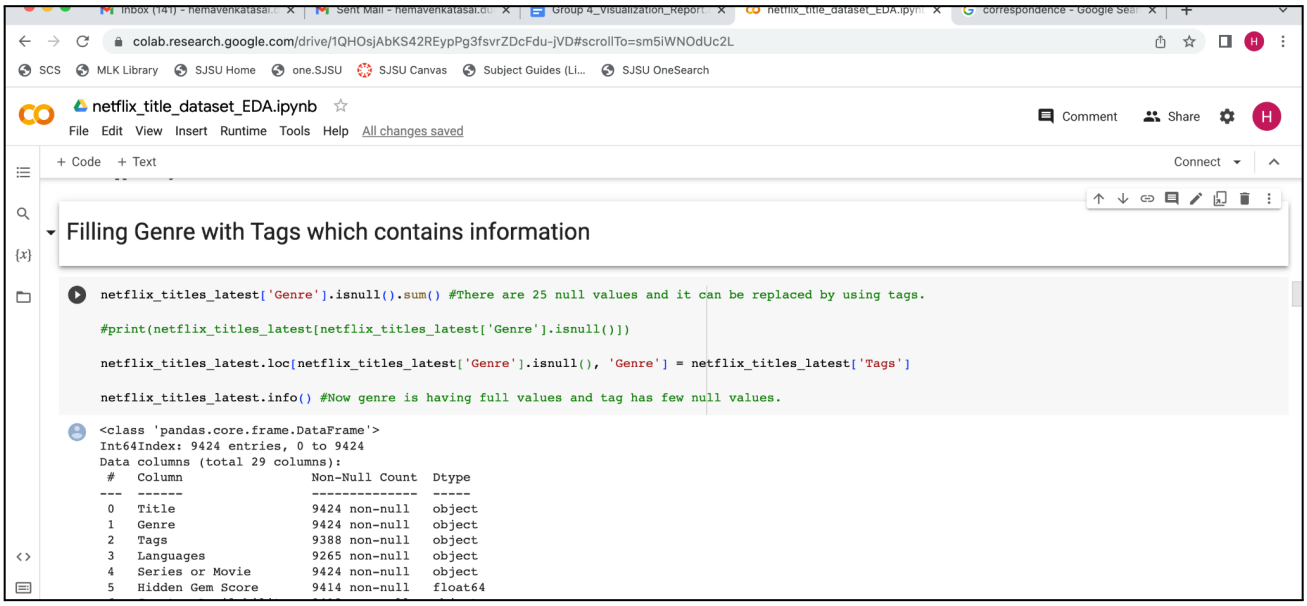


Step 3: Some columns don't have required information and contain null values. So,

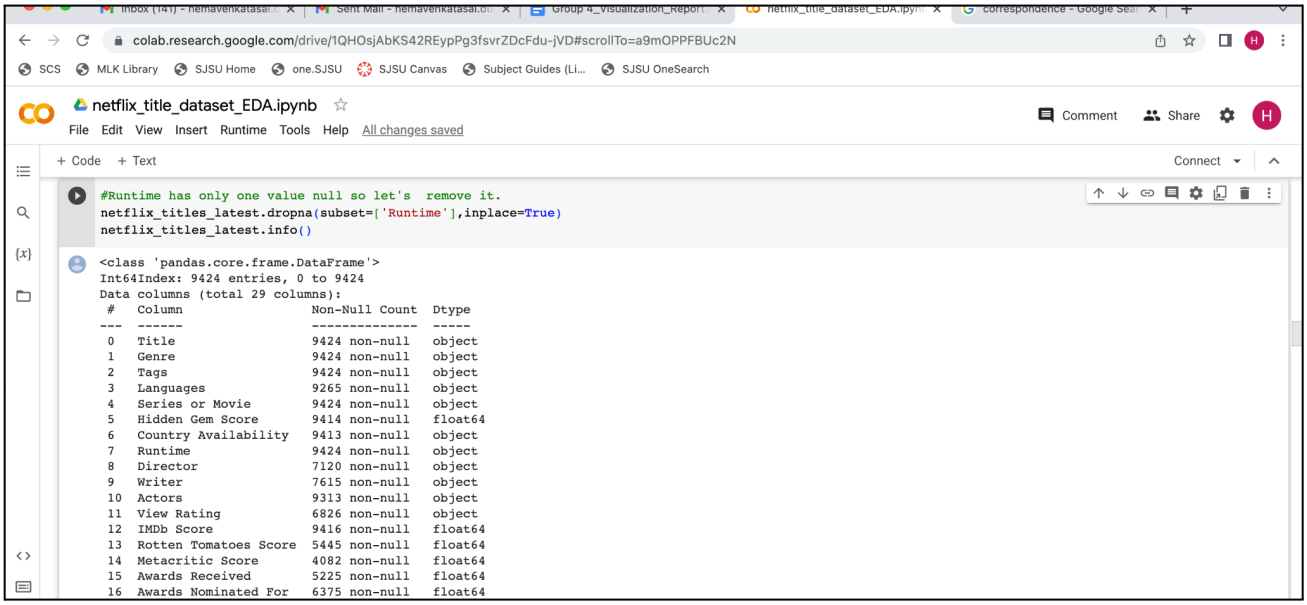
Those are handled by filling the null values with default data and filled with

corresponding values using correlation matrix.

i) Filling Null values where the data is not present in fewer columns.



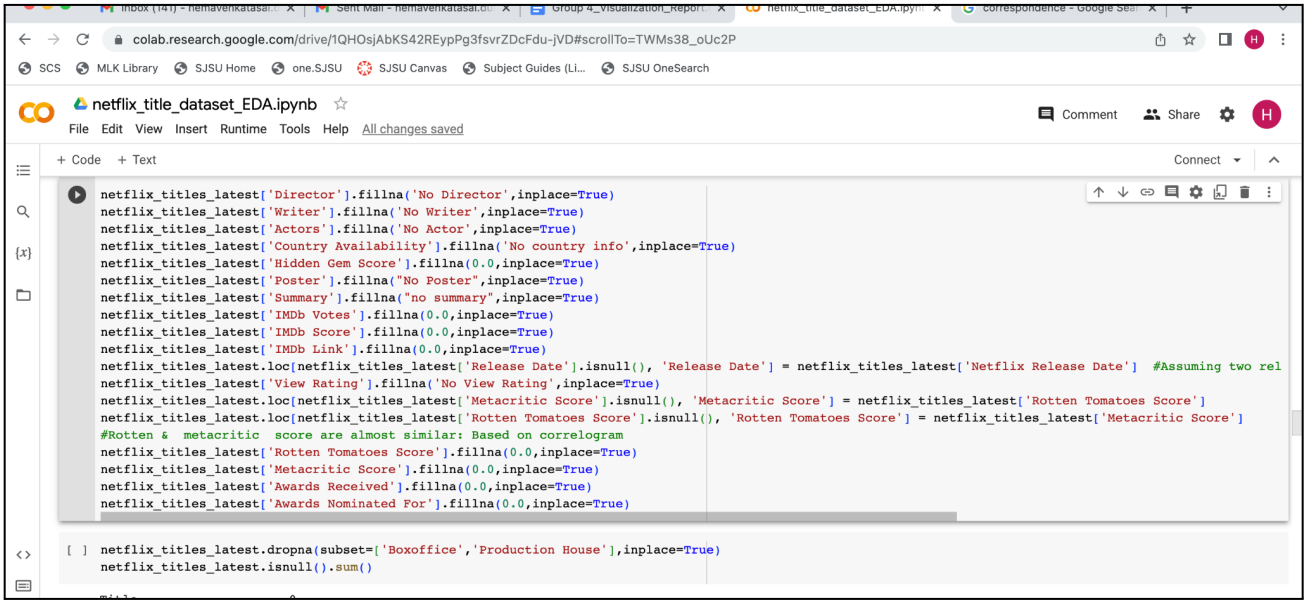
iii) Dropping the columns which are in less rows.



```
#Runtime has only one value null so let's  remove it.
netflix_titles_latest.dropna(subset=['Runtime'],inplace=True)
netflix_titles_latest.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9424 entries, 0 to 9424
Data columns (total 29 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Title                  9424 non-null   object
 1   Genre                  9424 non-null   object
 2   Tags                   9424 non-null   object
 3   Languages              9265 non-null   object
 4   Series or Movie        9424 non-null   object
 5   Hidden Gem Score       9414 non-null   float64
 6   Country Availability   9413 non-null   object
 7   Runtime                9424 non-null   object
 8   Director               7120 non-null   object
 9   Writer                 7615 non-null   object
 10  Actors                 9313 non-null   object
 11  View Rating            6826 non-null   object
 12  IMDb Score             9416 non-null   float64
 13  Rotten Tomatoes Score  5445 non-null   float64
 14  Metacritic Score       4082 non-null   float64
 15  Awards Received        5225 non-null   float64
 16  Awards Nominated For   6375 non-null   float64
```

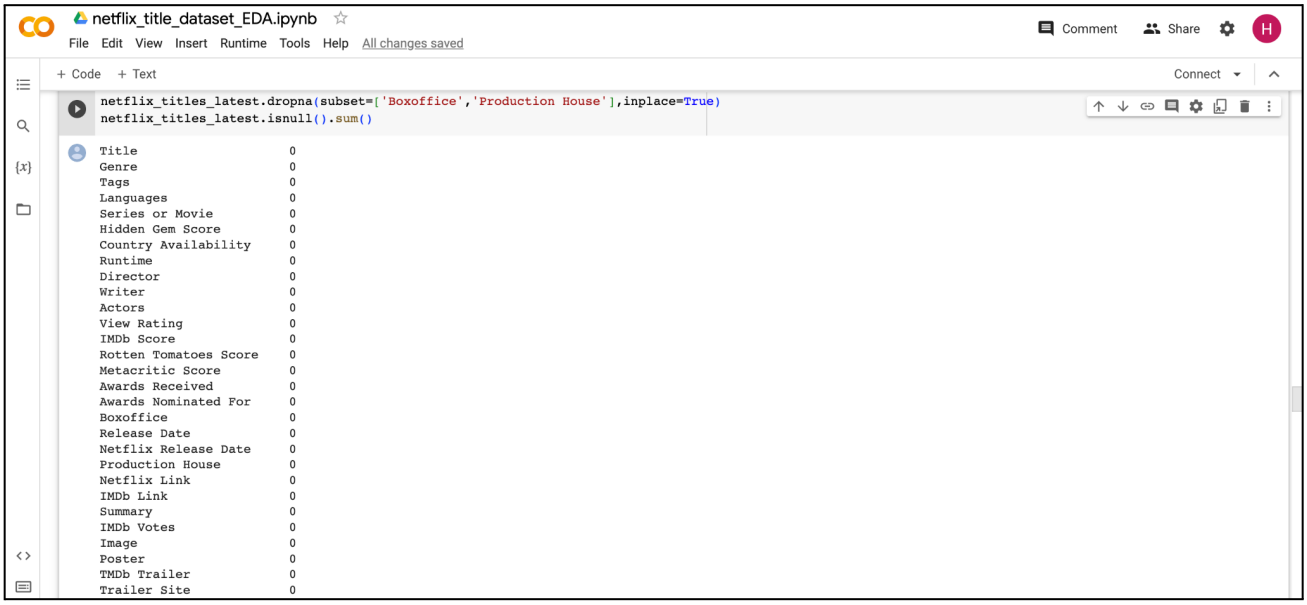
iv) Finally filling the columns with default values.



```
netflix_titles_latest['Director'].fillna('No Director',inplace=True)
netflix_titles_latest['Writer'].fillna('No Writer',inplace=True)
netflix_titles_latest['Actors'].fillna('No Actor',inplace=True)
netflix_titles_latest['Country Availability'].fillna('No country info',inplace=True)
netflix_titles_latest['Hidden Gem Score'].fillna(0.0,inplace=True)
netflix_titles_latest['Poster'].fillna("No Poster",inplace=True)
netflix_titles_latest['Summary'].fillna("no summary",inplace=True)
netflix_titles_latest['IMDb Votes'].fillna(0.0,inplace=True)
netflix_titles_latest['IMDb Score'].fillna(0.0,inplace=True)
netflix_titles_latest['IMDb Link'].fillna(0.0,inplace=True)
netflix_titles_latest.loc[netflix_titles_latest['Release Date'].isnull(), 'Release Date'] = netflix_titles_latest['Netflix Release Date']  #Assuming two rel
netflix_titles_latest['View Rating'].fillna('No View Rating',inplace=True)
netflix_titles_latest.loc[netflix_titles_latest['Metacritic Score'].isnull(), 'Metacritic Score'] = netflix_titles_latest['Rotten Tomatoes Score']
netflix_titles_latest.loc[netflix_titles_latest['Rotten Tomatoes Score'].isnull(), 'Rotten Tomatoes Score'] = netflix_titles_latest['Metacritic Score']
#Rotten &  metacritic  score are almost similar: Based on correlogram
netflix_titles_latest['Rotten Tomatoes Score'].fillna(0.0,inplace=True)
netflix_titles_latest['Metacritic Score'].fillna(0.0,inplace=True)
netflix_titles_latest['Awards Received'].fillna(0.0,inplace=True)
netflix_titles_latest['Awards Nominated For'].fillna(0.0,inplace=True)
```

```
[ ] netflix_titles_latest.dropna(subset=['Boxoffice','Production House'],inplace=True)
    netflix_titles_latest.isnull().sum()
```

Step 4: Converting the date column with required columns.



After the cleaning of all the columns. We got all the data without any null or missing values and now we are ready for finding the insights from the columns.

8. VISUALIZATIONS

Visualizations were done using several technologies such as tableau and python libraries. These methods were performed on two different datasets, one out of which is survey data and the others is the data that was scraped. Different insights were produced by visualizing the data with plots as shown below.

« The data has undergone preprocessing procedures including tokenization, stop word removal, and maintaining just alphabetical words, as seen by the word clouds being constructed based on the 'Preprocessed_Description' column of the 'df' DataFrame. This preprocessing aids in concentrating on significant terms and raises the standard of the word clouds.

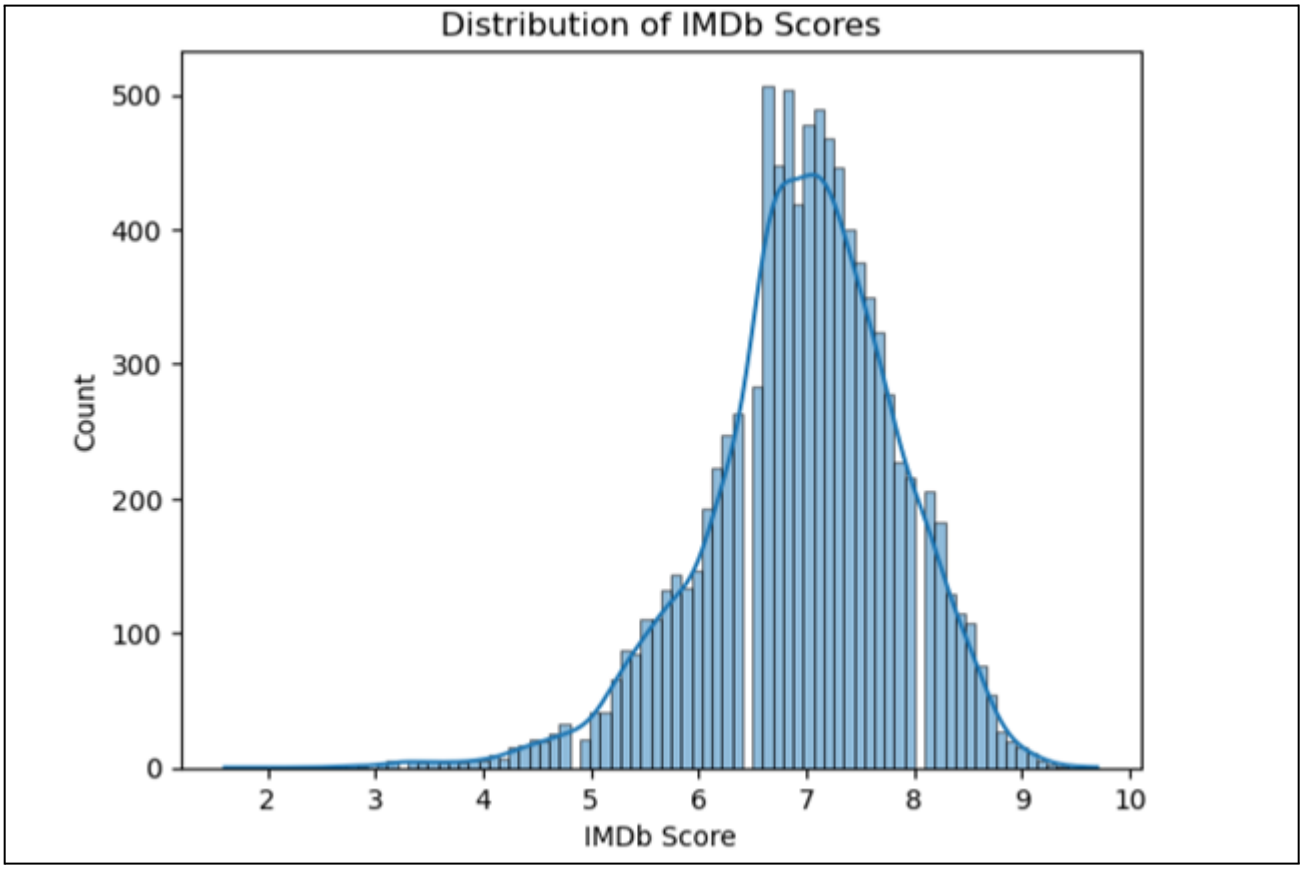
« Visual exploration of the terms most frequently used to describe each genre is possible. This might provide light on the themes and subjects that are common across many genres.

« The height and prominence of terms in word clouds indicates how frequently they appear in

genre-specific movie or television program descriptions. Words that are bigger and more noticeable imply a higher frequency.

« Word clouds that emphasize particular or unusual terms connected to each genre might aid in genre differentiation. This can shed light on the traits that distinguish distinct genres.

«



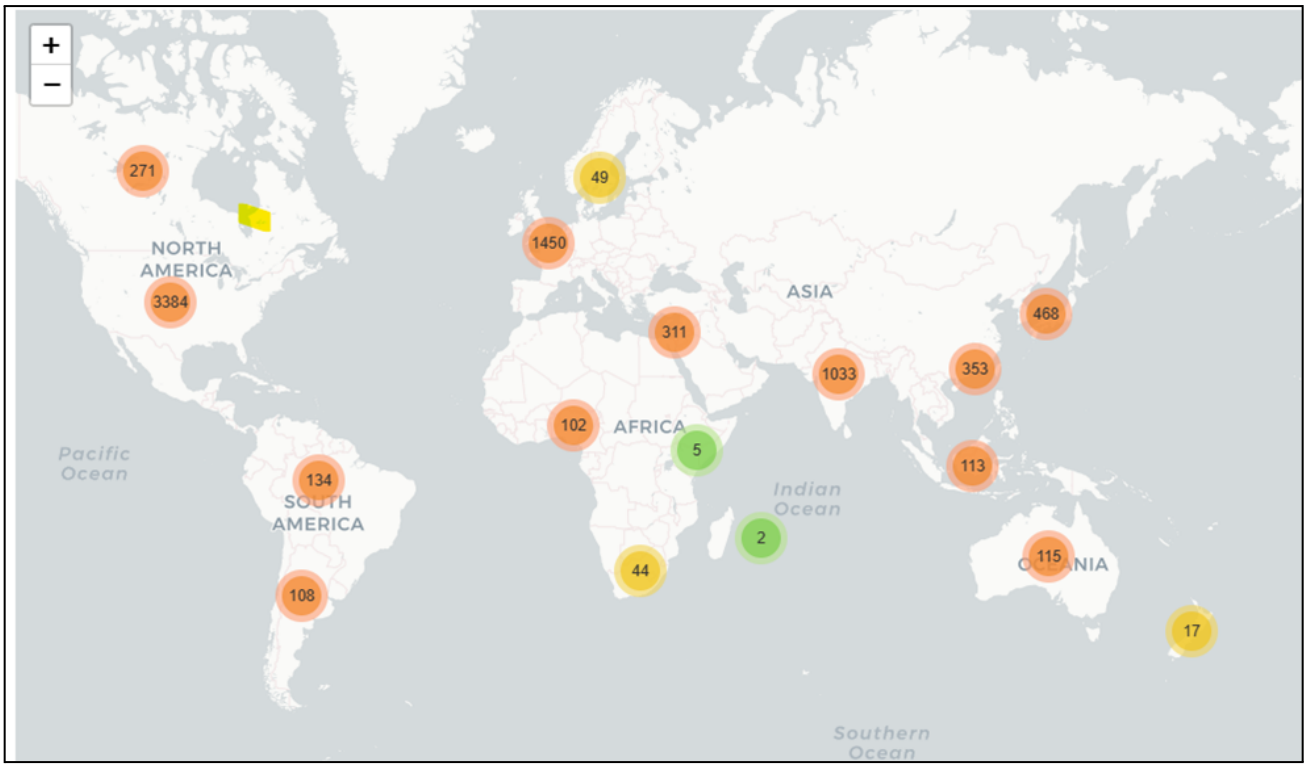
« IMDb scores have a somewhat typical distribution, peaking at around 6-7 on the scale.

« IMDb ratings for the majority of the Netflix material vary from 6 to 8, suggesting a generally acceptable level of quality.

« The right side of the distribution has a long tail, indicating that a small number of great movies or TV series have received very high IMDb ratings.

« The histogram reveals that titles with extremely low IMDb scores (below 4) are less prevalent than those with higher rankings. This indicates that Netflix has a range of material that is frequently

enjoyed by viewers.



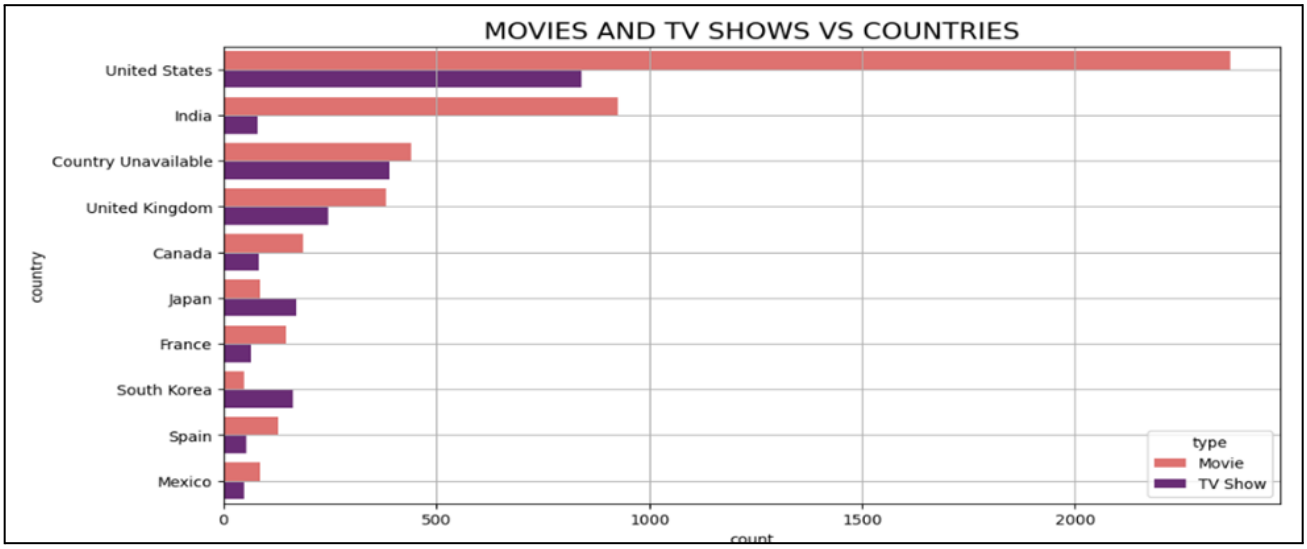
★ The map indicates where Netflix's movies and TV series are located in relation to one another. The size of each marker, which stands in for a different nation, gives a clue as to how many films and TV episodes are offered there.

★ Disparities in material Availability: The accessibility of Netflix material varies by nation. It's possible that some nations have more movies and TV series than others. By contrasting the sizes of the map markers, it may be concluded that this.

★ Popular locations: The markers on the map may be used to determine which locations have a higher concentration of Netflix content. Larger marker clusters may denote areas with a greater concentration of films and television programs.

★ Incomplete or missing data may exist for some nations. This is evident when the popup text for markers is blank or has insufficient information.

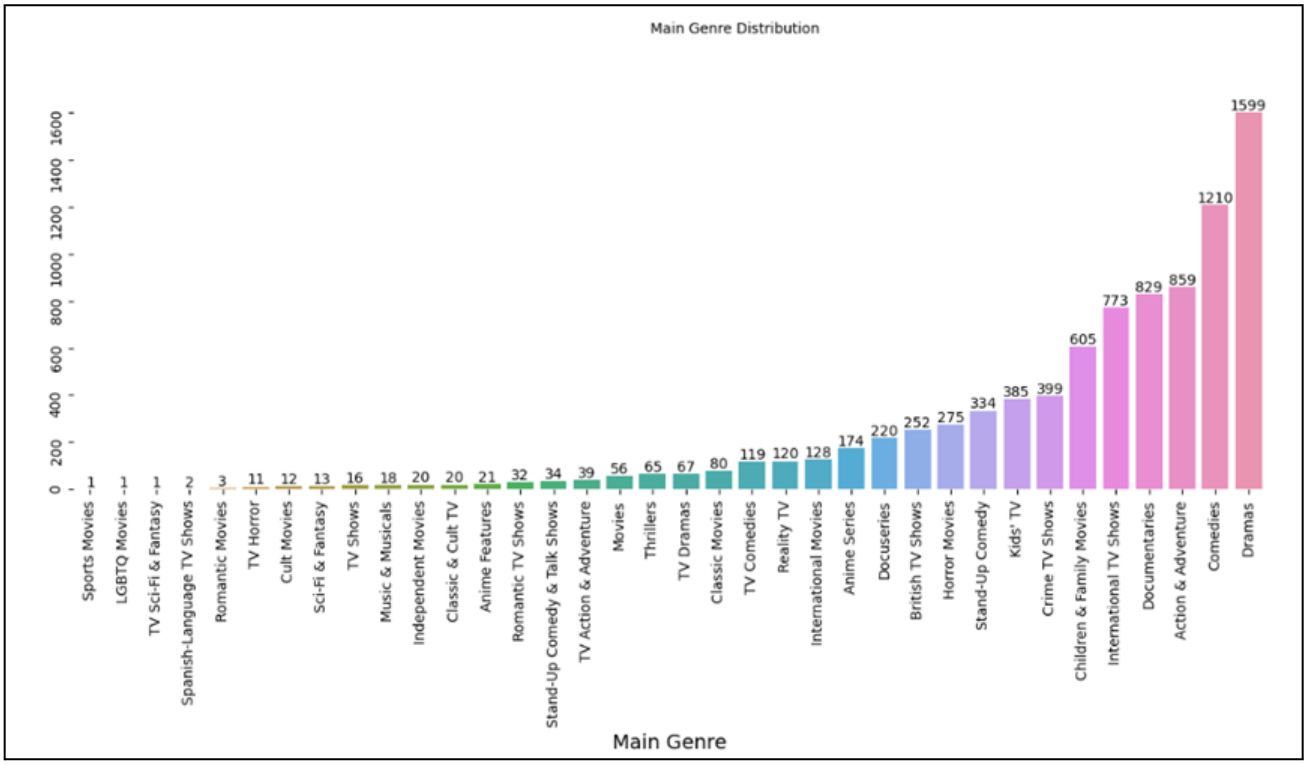
★ Netflix's dataset and the quality of the geolocation data utilized therein determine the accuracy of the content counts and geographic locations. Data errors or discrepancies might affect the visual representation and the conclusions drawn from it.



★ The countplot visulaization shows the lists the top 10 nations for Netflix content production, broken down by kind (movies and TV series).

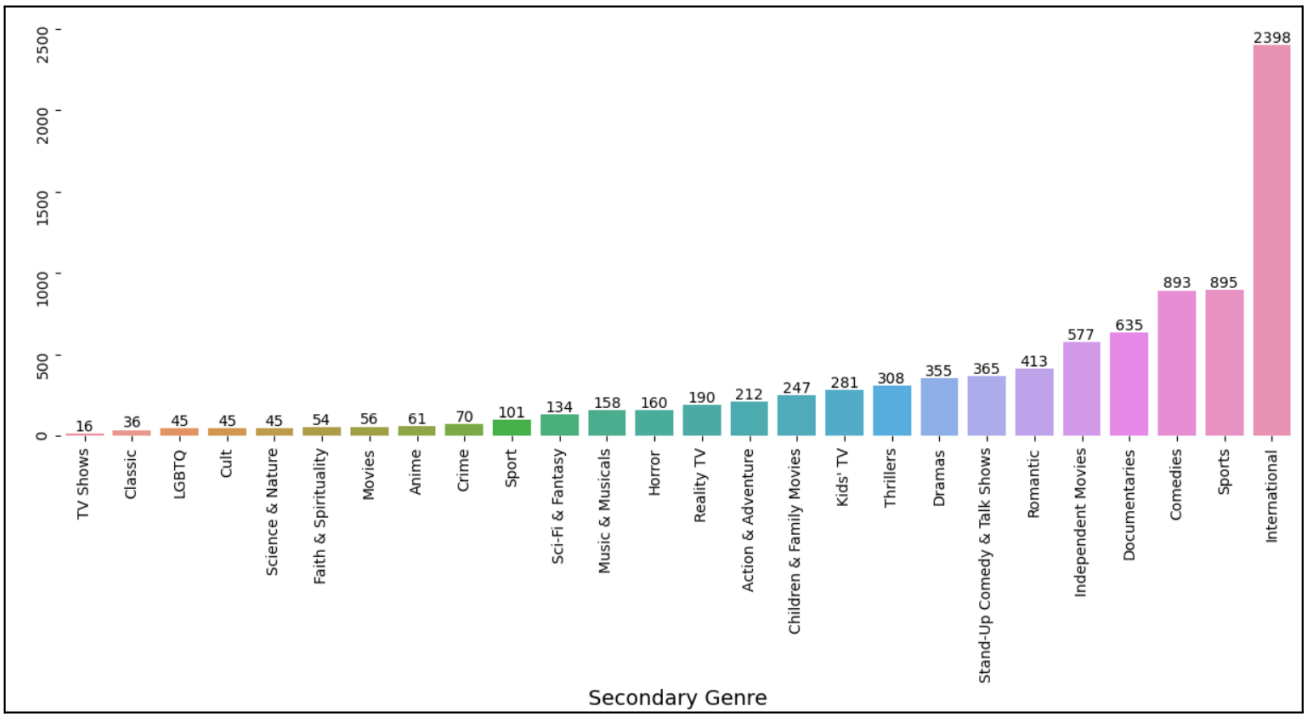
★ The countplot offers a visual comparison of the quantity of motion pictures and television programs generated in each nation. It illustrates that, in comparison to the other top 10 nations, the United States generates a disproportionately bigger volume of films and television series.

★ In terms of content creation, India is well-represented on Netflix, especially in the TV program category.



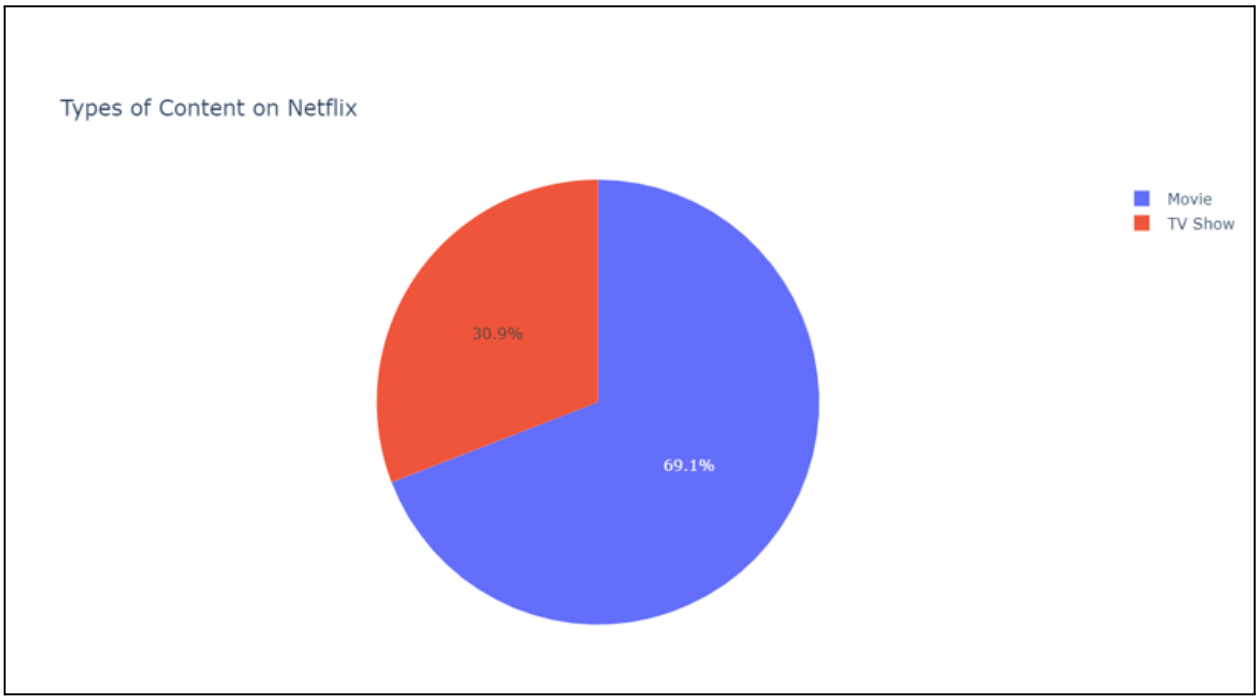
Distribution of Main Genres:

- ★ The storyline offers a visual depiction of how the content is split up across the several Main Genres. The height of each bar, which represents a primary genre, shows how much content falls within that category.
- ★ The count plot makes it simple to see which genres have the highest and lowest frequencies among the major ones. Shorter bars reflect genres with less content items, whereas taller bars show genres with more content items.
- ★ The most well-liked primary genres in the dataset may be found by evaluating the highest bars in the plot. These genres include more content than others, which may indicate more interest in or output of content in particular genres.
- ★ Contrarily, genres with extremely small bars denote unusual genres with a reduced number of content pieces. These subgenres in the dataset can be specialized or uncommon.



Distribution of Secondary Genres:

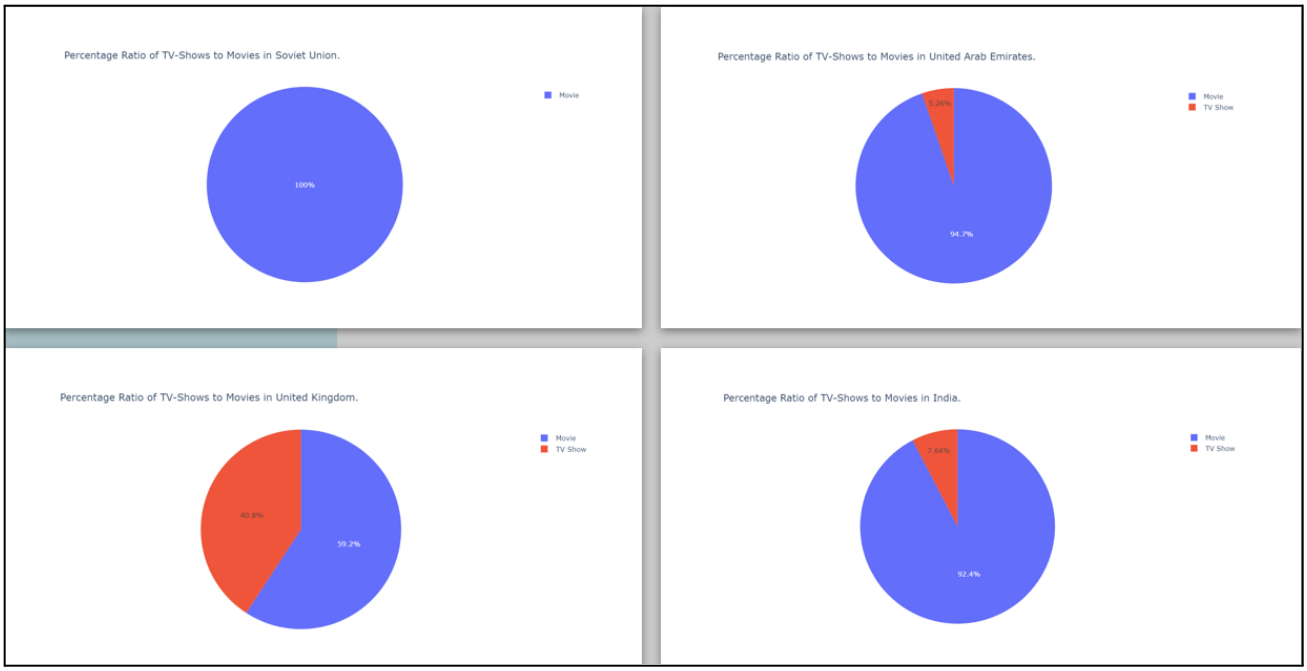
- ★ This count plot displays the distribution of secondary genres in the dataset, much as the primary genre distribution plot.
- ★ The secondary genres are shown on the x-axis, while the count is shown on the y-axis.



★ The above piechart gives an overview of the dataset.

★ Despite Netflix's popularity for TV shows, its movies are more prevalent.

★ TV shows make up 30% of the dataset, while movies make up 70%.

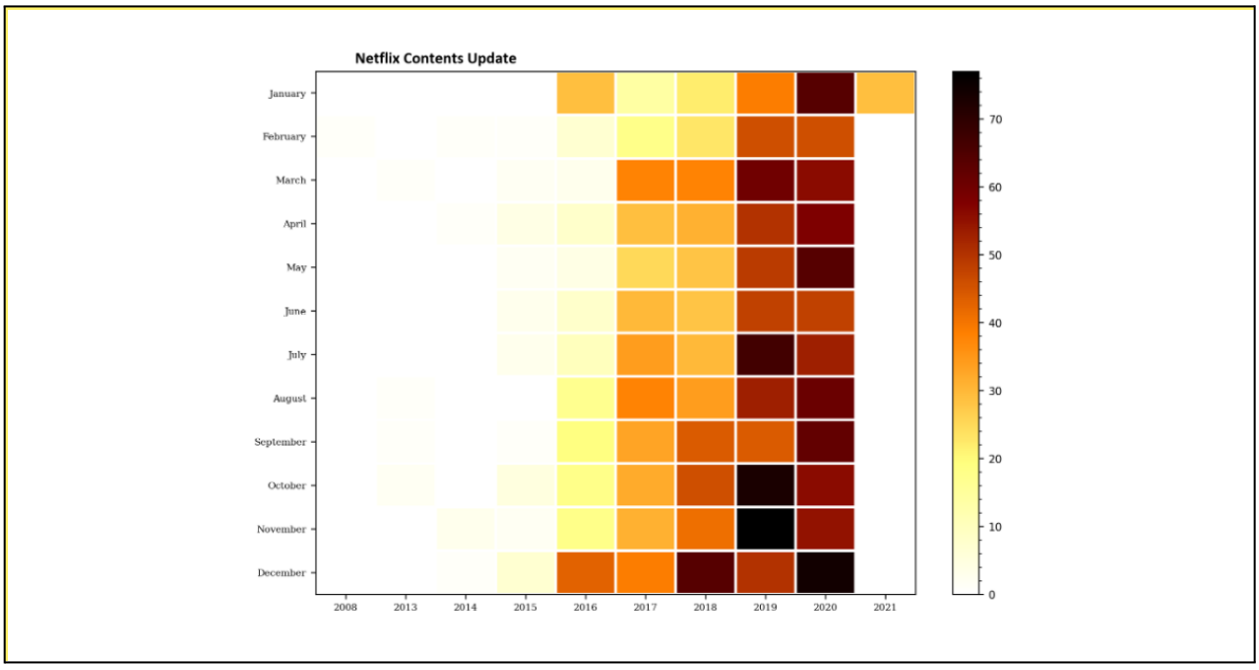


★ In the piechart dashboard above, TV shows and movies are compared across four different countries.

★ Compared to other countries, the United Kingdom has the highest ratio of TV shows to movies.

★ The Soviet Union, however, has only movies and no TV shows. There is a place for Netflix to bring out innovative television show content that can be marketed in countries like these.

★ In India and UAE, the percentage of TV show content is minimal. Television shows seem to be emerging in these countries while movies dominate.



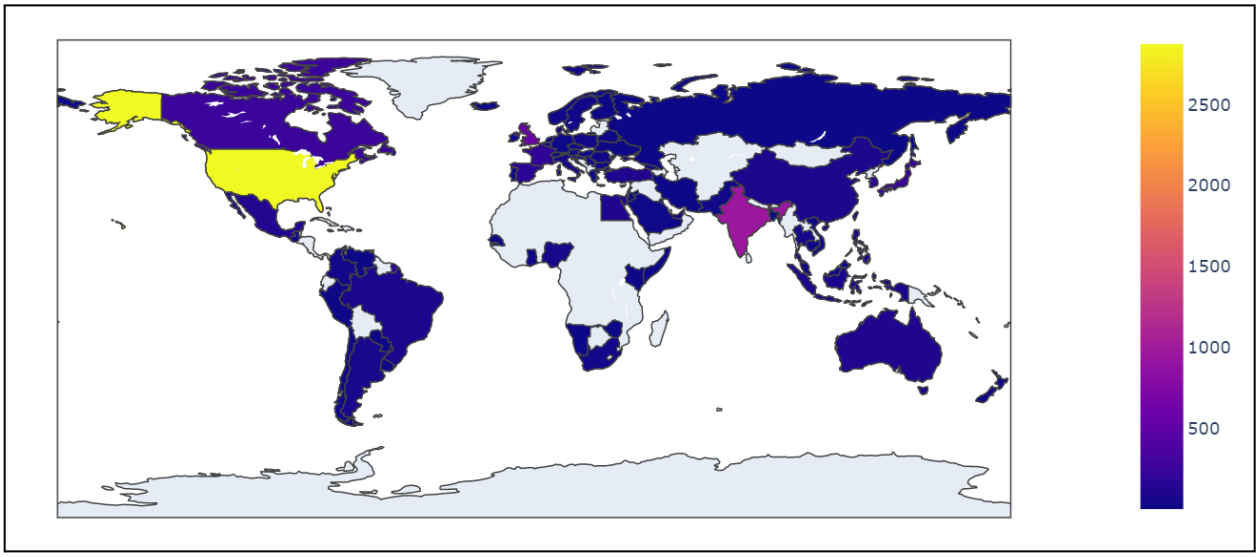
★ This heatmap shows the number of movies released each month from 2008 to 2021.

★ Movies were released in greater numbers in the months of October to December in the earlier years (2016-2018).

★ It could be due to winter and Christmas vacations, when people are more likely to watch movies.

★ But as the years passed on, especially in the years 2019 and 2020, the color gradient is constantly on the higher scale.

★ Due to the COVID pandemic, remote work had become the norm, and it is apparent that the month is no longer a factor when deciding when to release a movie.



★ In the above world map, you can see how much Netflix content each country produces per year.

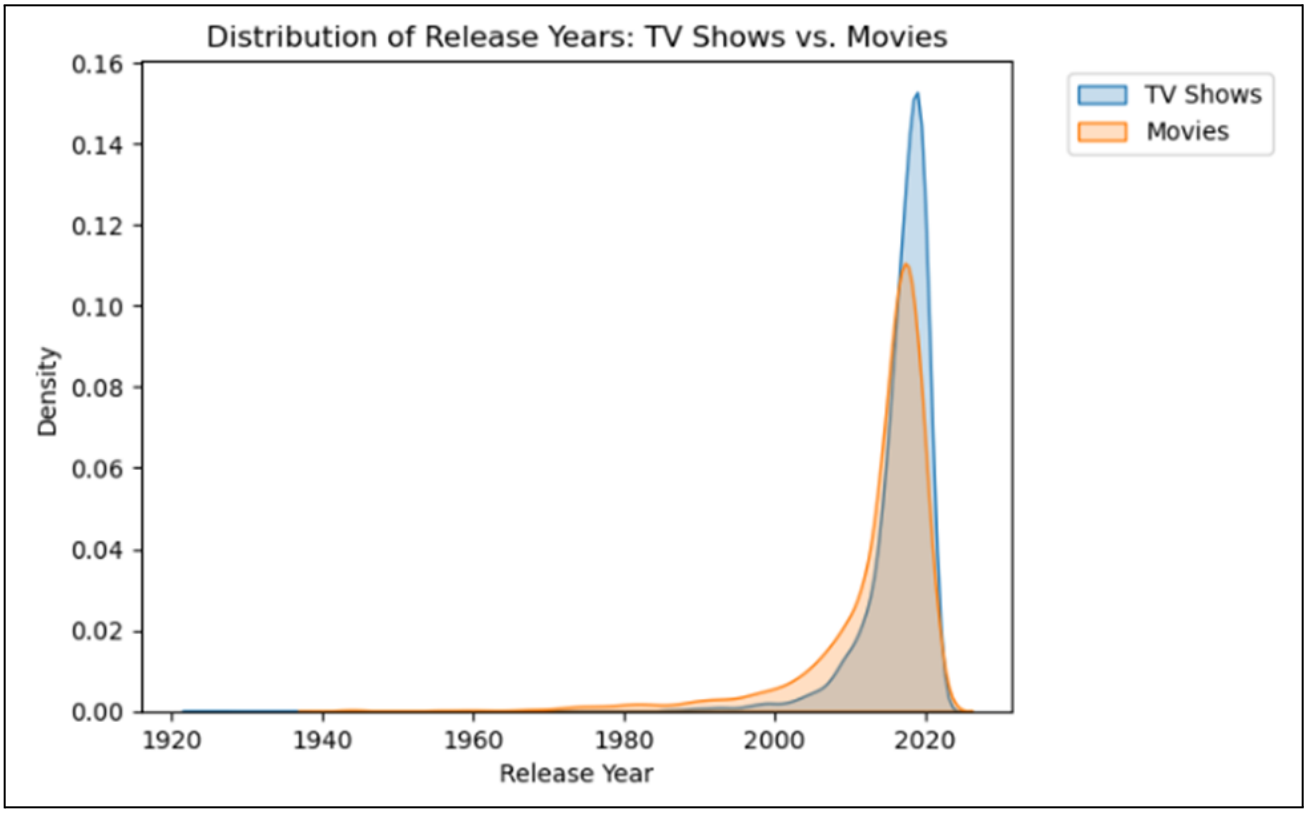
★ The scale on the left indicates the number of contents produced.

★ Around 2700 movies and TV shows are produced annually in the United States.

★ Following America, INDIA produces around 1500 movies and television shows a year.

★ Other major countries like Australia, UK, France etc seem to be on a similar scale.

★ It would be a great insight for the platform if it could focus more on which areas of the globe need improvement for it to expand.



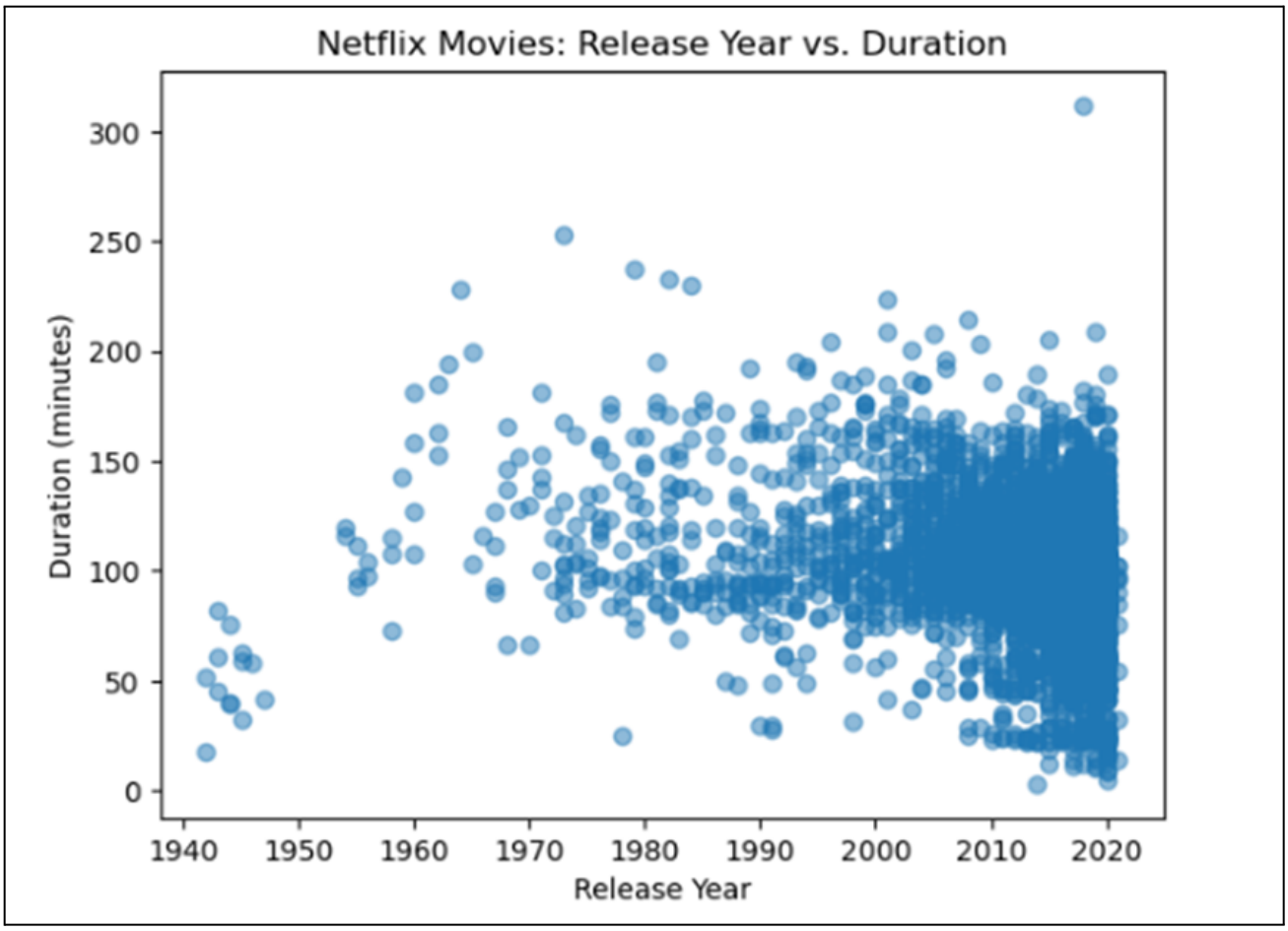
The above density plot describes the distribution of Netflix content over Period of Years. Color is used to distinguish both the movies and TV shows in the graph.

Observations:

- ★ · During the initial years, there is very less content that has been generated in Netflix.
- ★ · From 2000 onwards, Netflix has made massive releases and updated its content.
- ★ · Over the period of years, number of TV shows released is far more than Movies released.

Insights:

- ★ · Netflix might have observed some reason like people watching more TV shows than Movies, due to the easy going.



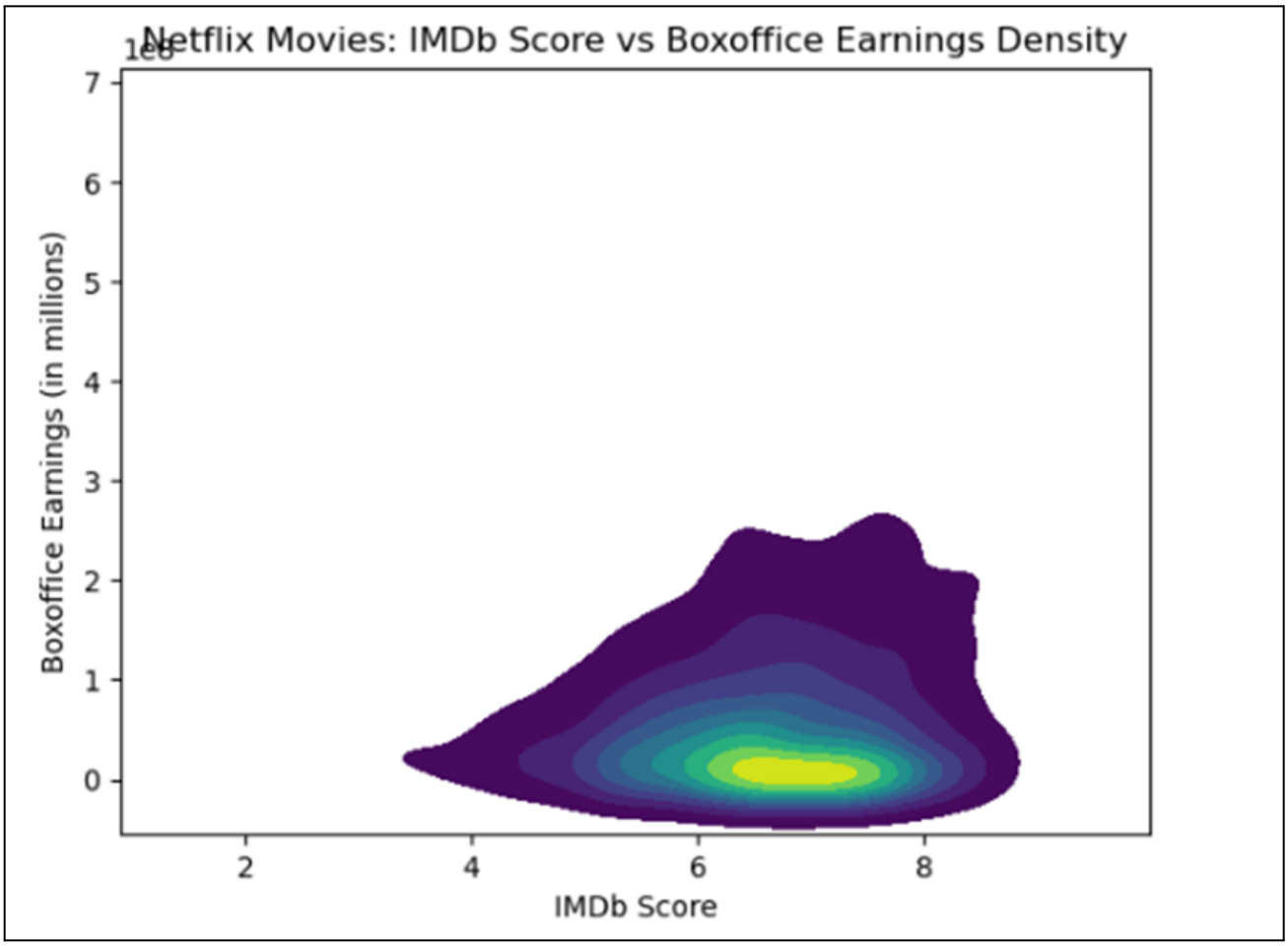
Above scatter plot shows the relationship between Release year and Duration of the Movies.

Observations:

- ★ · Majority of the movies released have the duration of 75-120 minutes.
- ★ · More number of movies are released after 2000.
- ★ · If we move towards more recent years, there appears to be a higher concentration of movies with shorter durations.

Insights:

- ★ · As the human watchable duration is between the above range, Netflix also concentrated on releasing movies of the above time range.
- ★ · While certain release years may exhibit a higher density of movies within specific duration ranges, there is no consistent trend of movie durations consistently increasing or decreasing over time.



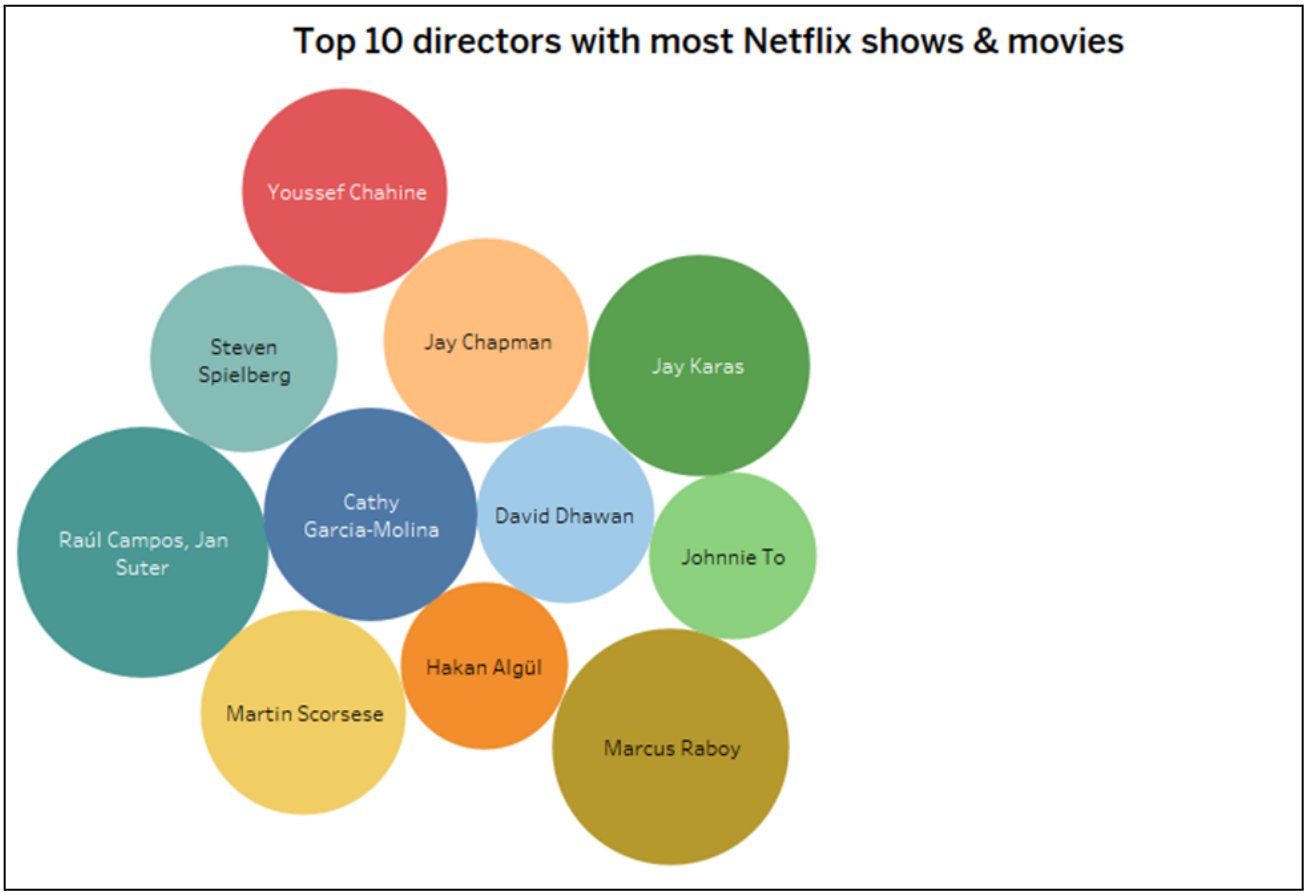
The density plot shows how the IMDB score of a movie impacts the box office earnings.

Observations:

- ★ Majority of Netflix movies have IMDb scores between 6 and 8.
- ★ This suggests that Netflix tends to offer movies with relatively high ratings, as indicated by the peak in density within this range.
- ★ The density plot highlights a higher concentration of movies with moderate IMDb scores (around 6 to 7) and moderate Box office.

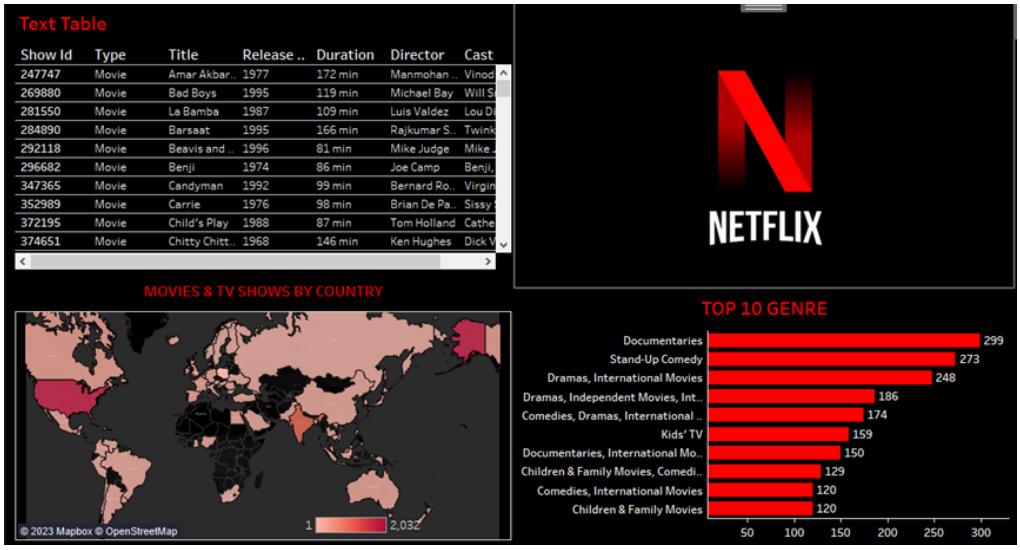
Insights:

- ★ The density plot also shows a positive correlation between IMDb scores and Boxoffice earnings.
- ★ As the IMDb scores increase, the density of movies with higher Boxoffice earnings also increases. This indicates that movies with better ratings tend to have higher box office performance.



The above bubble chart shows the top 10 directors who made the most number of TV shows and movies.

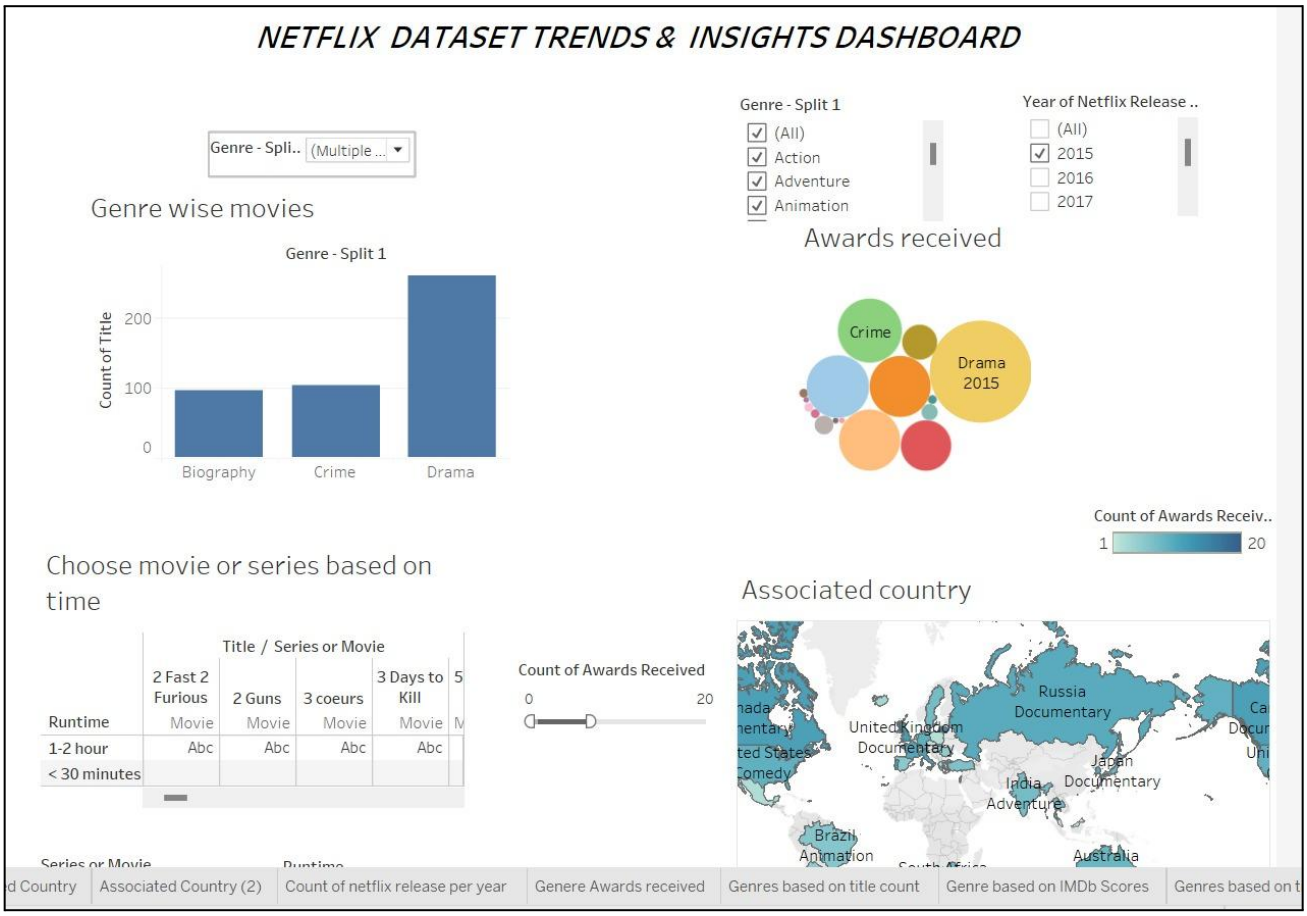
- ★ · Color is used to distinguish each director and size indicates the number of TV shows and movies that they have directed.
- ★ · Raul Campos, Jan Sutar stood at the first place with a huge majority followed by Marcus Raboy.



Above is a simple dashboard which was generated by collating three different types of charts i.e text table, geo map, horizontal bar charts and Netflix logo.

- ★ · The text table gives in detail information like released year, title, cast, directors about each movie and tv show that has been released on Netflix.
- ★ · Geo map gives the information above the number of movies and TV shows released in each country and color gradient is used to indicate the concentration of content. Darker the gradient, larger the content.
- ★ · Horizontal bar chart shows the top 10 genres of Netflix where Documentaries stood at first place with 299 followed by Stand-Up comedy.

DASHBOARD:



Finally, The dashboard, which contains Netflix trends and insights. Basically, Here we can see the count of genres for the corresponding years which ranges from 2015- 2021 and also we can check the awards received for particular genres which helps the stakeholders to improve their content for next year and also by using bar graphs for genres. The stakeholders can check the required information from the graph to improve their content which is very less in their domain. Other than this, at the left lower, We have a movie recommendation which is in initial development and it helps us to choose the series or movies based on our runtime that could range from 30 min, 1-2 hours and greater than 2 hours. At last, We have a map to check the associated country which has the awards for the genre for particular years or year.

9. CONCLUSION

Also, we were able to figure out the taste in movies and Tv shows of various countries, which would tell the platform which aspects to focus more on.

In conclusion, the study of the Netflix dataset has shed light on the functionality and user preferences of the platform. The results show that there are a number of regions and nations that provide room for development. In order to improve its offers and broaden its worldwide reach, Netflix may strategically target these regions by identifying them.

Additionally, in order to adapt its programming to the various inclinations of its consumers, Netflix must have a thorough awareness of global film and television tastes. Netflix is now able to devote resources and select material that appeals to particular audiences since the analysis has revealed the genres and categories that are popular in distinct geographic areas. This information can aid the platform's decision-making when purchasing and creating new content.

10. FUTURE SCOPE

Analyzing user data to anticipate and offer content based on personal interests is a key step in putting machine learning approaches for movie/TV show recommendations into practice. This covers the application of hybrid models, content-based filtering, and collaborative filtering. Web scraping also makes it possible to gather up-to-date information on impending events, like release dates, cast information, and narrative summaries. By providing information on forthcoming material, this data improves recommendation algorithms and enables personalized suggestions even before the item is released. It also increases the accuracy of content suggestions, recognizes new trends, and helps with content purchasing choices. These strategies ultimately improve customer satisfaction, engagement, and competitiveness in the streaming sector.

11. REFERENCES

- https://www.researchgate.net/publication/360856267_Netflix_Recommendation_System_based_on_TF-IDF_and_Cosine_Similarity_Algorithms by Mohamed Chiny, Marouane Chiha , Omar Benchare and Younes Chiha.
- https://www.researchgate.net/publication/354719521_Exploratory_and_Sentiment_Analysis_of_Netflix_Data.
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3473148
- Krishna Prasad, K. & Aithal, P. S. (2018). A Study on Multifactor Authentication Model Using Fingerprint Hash Code, Password and OTP. International Journal of Advanced Trends in Engineering and Technology, 3(1), 1-11. DOI: http://doi.org/10.5281/zenodo.1135255.
- Pazzani M.J., Billsus D. (2007) Content-Based Recommendation Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg
- Netflix. How Netflix Recommendation System Works. Retrieved 09/08/2019, from https://help.netflix.com/en/node/100639

DEPLOYMENT

Youtube Link: https://www.youtube.com/watch?v=Aswk5EcKlBA

Github Link: https://github.com/pavankumarreddykasarla/225_Project_Group4