**DISNEY HOTSTAR ANALYSIS**

**A MINI PROJECT REPORT**

Submitted by

RAHUL K (22ADR084)

NEELESHWAR K (22ADR074)

RANJITH KUMAR P (22ADR086)

for

22ADF01 DATA ANALYSIS

DEPARTMENT OF ARTIFICIAL INTELLIGENCE



KONGU ENGINEERING COLLEGE

(Autonomous)

PERUNDURAI ERODE – 638060

NOVEMBER 2024

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE KONGU ENGINEERING COLLEGE**

**(Autonomous)**

**PERUNDURAI ERODE – 638 060**

Department of Artificial Intelligence

**22ADF01 – Data Analysis Project Report**

Signature of course in-charge Signature of the HOD

Submitted for the continuous Assessment viva voice examination held on

EXAMINER I EXAMINER II

# ABSTRACT

This Power BI analysis focuses on Disney+ Hotstar, a leading streaming platform offering diverse content, including movies, series, and live events. The analysis aims to uncover insights into the platform’s content catalog, user preferences, and trends across various dimensions such as genres, age ratings, runtime, and content types.

By leveraging a dataset containing fields like **title, genre, year, age rating, running time, seasons, and episodes**, the analysis highlights the distribution of content types (movies vs. series), the prevalence of popular genres, and patterns in age-specific content offerings. Additionally, it examines the platform's evolution over time, identifying trends in content production and audience engagement.

This study provides valuable insights for understanding Disney+ Hotstar's strategy in catering to a global audience. It also offers actionable intelligence for stakeholders looking to enhance user engagement, optimize content offerings, and maintain a competitive edge in the dynamic streaming industry.

|  |  |  |
| --- | --- | --- |
| **CHAPTER**  **No.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | 3 |
| **1.** | **INTRODUCTION** |  |
|  | 1.1 INTRODUCTION | 5 |
|  | 1.2 DATA COLLECTION | 5 |
|  | 1.3 PROBLEM STATEMENT | 6 |
|  | 1.4 BUSINESS OBJECTIVE | 7 |
| **2.** | **DATA PREPARATION AND MODELING** |  |
|  | 2.1 DATA PREPARATION | 8 |
|  | 2.2 DATA MODELLING | 17 |
| **3.** | **DATA ANALYSIS AND INTERPRETATION** |  |
|  | 3.1 DATA ANALYSIS | 20 |
|  | 3.2 PUBLISHING DASBOARDS | 22 |
|  | 3.3 INFERENCE | 23 |
| **4.** | **CONCLUSION** | 40 |
| **5.** | **REFERENCES** | 41 |

# TABLE OF CONTENTS

**CHAPTER 1**

**INTRODUCTION**

# 1.1 INTRODUCTION

Disney+ Hotstar is a leading video-on-demand streaming platform offering a diverse library of movies, TV shows, sports, and original series. Catering to a global audience, it features content from Disney, Marvel, Pixar, Star Wars, and National Geographic, along with exclusive regional offerings.

With features like multilingual support, multiple subscription tiers, and live sports streaming, Disney+ Hotstar appeals to a broad range of viewers. It has transformed entertainment consumption by emphasizing accessibility, affordability, and personalized recommendations powered by data-driven insights.

# 1.2 DATA COLLECTION

The data used in this Disney+ Hotstar analysis was sourced externally from Kaggle, a popular platform for datasets and machine learning. Data collection is a critical process that involves gathering and compiling information from reliable sources to identify trends, patterns, and insights.

For this analysis, the dataset includes key attributes such as **title, genre, year, age rating, running time, seasons, episodes, and type** (movie or series). This structured data allows for comprehensive exploration of content distribution, user preferences, and trends.

By leveraging an external source like Kaggle, the analysis ensures access to a well-organized and diverse dataset, enabling accurate and actionable insights into Disney+ Hotstar's offerings and audience engagement.

DATASET:

The Disney+ Hotstar dataset, sourced from Kaggle, provides comprehensive information about the platform's content library, enabling detailed analysis of its offerings. The dataset contains variables such as **hotstar\_id**, a unique identifier for each content item; **title**, representing the name of the movie or series; and **description**, providing a brief synopsis of the content. Other variables include **genre**, which classifies the content into categories like Comedy, Action, or Drama; **year**, indicating the release year; and **age\_rating**, denoting the viewer age group suitability (e.g., PG, 18+). For series, the dataset includes information on **seasons** and **episodes**, while **running\_time** captures the duration in minutes. The **type** variable differentiates between movies and series. This dataset is valuable for analyzing content trends, understanding audience preferences, and exploring Disney+ Hotstar's approach to catering to a global and diverse audience.

**Source dataset**: [Kaggle - Disney+ Hotstar Dataset](https://www.kaggle.com/datasets/sanjanaalaham/disney-hotstar-dataset?resource=download)

Source: www.kaggle.com

The Disney+ Hotstar dataset provides valuable insights into the platform's vast and diverse content library. Beyond basic details such as **title**, **description**, and **genre**, the dataset enables deeper analysis into content evolution over time. For example, by examining the **year** of release and **age\_rating**, one can uncover how the platform’s content strategy has adapted to changing viewer demographics and preferences. Furthermore, **running\_time**, **seasons**, and **episodes** help identify trends in content length, whether viewers prefer short films, long-running series, or multi-season shows.

Another significant variable is **type**, which classifies content as either a **Movie** or a **Series**, offering insight into Disney+ Hotstar's investment in original series versus standalone films. Analyzing these variables helps identify key periods of growth in the platform’s series catalog, such as during the rise of binge-watching culture. This dataset can be used to assess the global reach of Disney+ Hotstar's content, examining how content is tailored for specific regional audiences while maintaining global appeal.

With such rich data, it’s possible to perform trend analysis, forecast future content releases, and evaluate how the platform’s offerings compare with competitors in the streaming space.

# 1.3 PROBLEM STATEMENT

The problem at hand is to analyze the content library of Disney+ Hotstar to identify trends, patterns, and strategies that can enhance the user experience and content offerings. The dataset contains a wealth of information about movies, TV shows, and series, including their genre, release year, age rating, running time, and more. This vast collection of data can provide valuable insights into the types of content that resonate most with audiences, how content distribution has evolved over time, and whether the platform is adapting to emerging trends like longer series or shorter films. By leveraging this dataset, it’s possible to understand Disney+ Hotstar’s content strategy and identify areas where the platform can improve its offerings to meet the demands of its global audience.

The challenge lies in extracting meaningful insights from the dataset that can guide strategic decisions for Disney+ Hotstar. This includes determining which genres are most popular in various regions, identifying which types of content—movies or series—are favored by different age groups, and recognizing patterns in content release over the years. Additionally, the platform’s shift towards producing original series and diversifying its content based on regional preferences needs to be explored. The goal is to use data analysis to drive recommendations for optimizing content acquisition, production, and curation strategies that align with current viewer preferences and future trends in the streaming industry.

# 1.4 BUSINESS OBJECTIVE

 **Optimize Content Strategy**: Analyze genre preferences, content types, and viewer demographics to guide Disney+ Hotstar in curating content that resonates with its global audience.

 **Enhance User Experience**: Use data insights to improve recommendations and content personalization, increasing user engagement and satisfaction on the platform.

 **Regional Content Targeting**: Identify trends in regional content preferences to tailor offerings and expand market reach effectively across diverse audiences.

 **Content Performance Analysis**: Evaluate the success of original series vs. movies and assess their impact on subscriber growth and retention, helping to refine future content investments.

**CHAPTER 2**

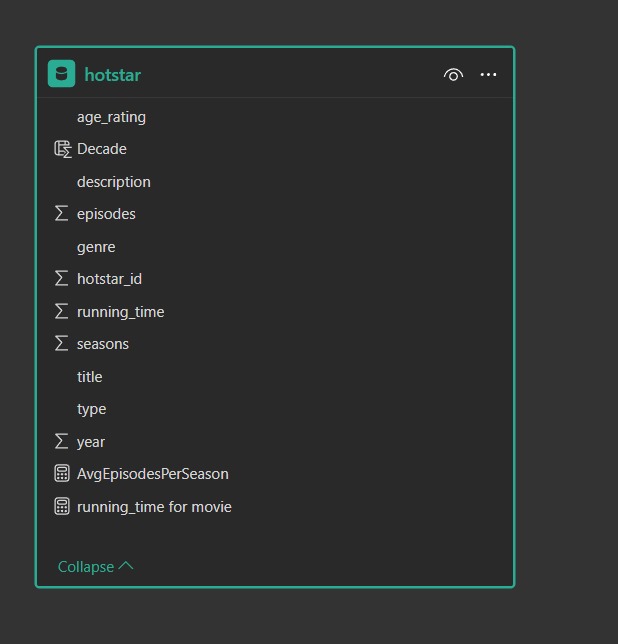
**DATA PREPARATION AND MODELLING**

# 2.1 DATA PREPARATION

The first step in preparing data for analysis is collecting relevant datasets. For Disney+ Hotstar, data can be sourced from multiple locations including internal databases, external datasets (e.g., Kaggle), and APIs. The dataset used for analysis includes key attributes like hotstar\_id, title, description, genre, release year, age rating, running time, seasons, episodes, and content type. Additional sources such as user ratings, subscription data, and regional preferences may also be collected to enrich the analysis.

##  DATA COLLECTION

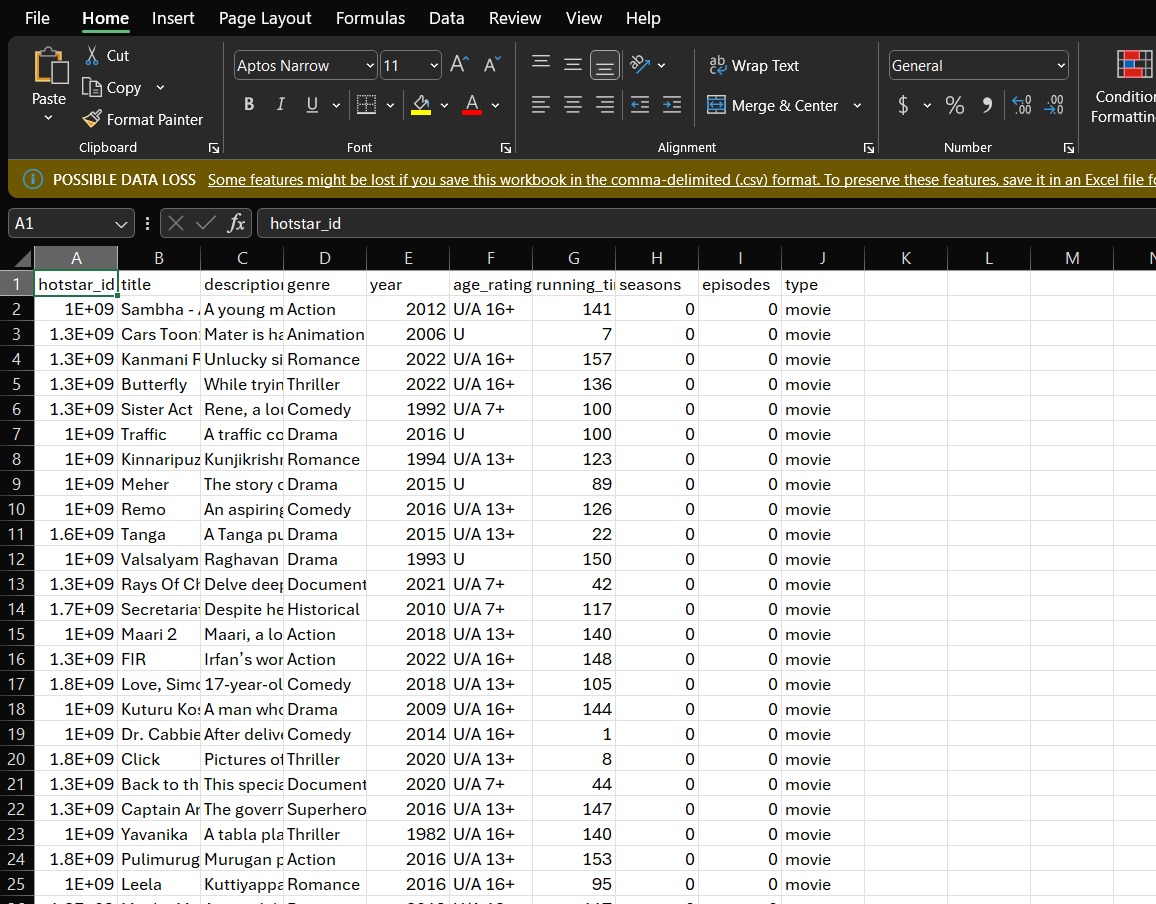
Often, data from different sources (internal and external) needs to be combined for a comprehensive analysis. For Disney+ Hotstar, combining datasets such as the content library dataset (movies, shows, etc.) with viewership data, user ratings, and subscription trends is essential. This integration allows for a better understanding of how content performance correlates with user engagement and regional preferences.



## Figure 2.1 Loading data

###  DATA DISCOVERY AND PROFILING

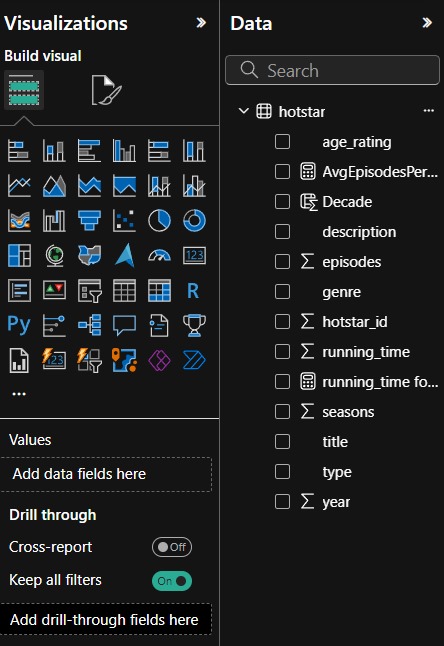
Data discovery and profiling are essential steps in understanding the structure, quality, and patterns within the dataset. For Disney+ Hotstar, data discovery involves exploring and identifying the various attributes and relationships within the dataset, such as content types (movies vs. series), genres, release years, age ratings, and other key factors like running time, number of seasons, and episodes. Profiling, on the other hand, involves analyzing the dataset to gain insights into its distribution and characteristics. For example, profiling may reveal which genres are most popular, the average running time of content, and trends in age ratings. It also helps identify potential data quality issues, such as missing values, duplicates, or inconsistencies in content categorization (e.g., misclassified genres or incorrect release years). By leveraging data profiling techniques, we can ensure that the dataset is clean, complete, and ready for further analysis, providing a solid foundation for decision-making regarding content strategy, user engagement, and platform growth.



## Figure 2.2 Data review

###  DATA STRUCTURING

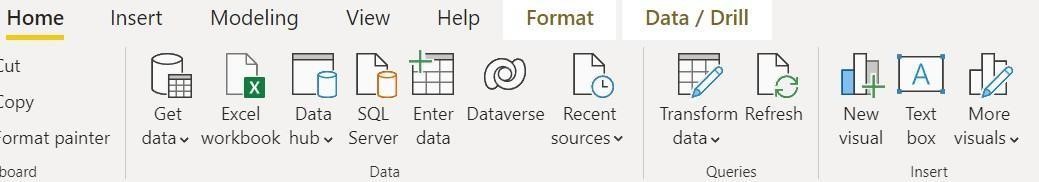
Data structuring involves organizing raw data into a clear, consistent, and usable format to facilitate analysis. For Disney+ Hotstar, structuring the dataset ensures that key information about the platform's content is organized in a way that makes it easy to query, analyze, and visualize. The dataset can be structured into tables or spreadsheets, with each row representing an individual piece of content, such as a movie or TV show. Columns would include attributes like **hotstar\_id** (a unique identifier), **title** (name of the content), **genre** (e.g., Action, Comedy, Drama), **release\_year**, **age\_rating** (e.g., PG, 18+), **running\_time** (duration in minutes), **seasons** (for series), **episodes** (for series), and **type** (Movie or Series). Each of these columns should be clearly defined with appropriate data types (e.g., text for titles, numeric for running time, categorical for genre and age rating). Ensuring that all content attributes are consistent and properly formatted is critical for analysis, as this structure allows for easy aggregation, comparison, and reporting. Additionally, this structure supports advanced querying to identify trends such as the most popular genres, average content length, or content release patterns over time. By organizing the data in this way, it becomes accessible and meaningful, providing insights that can drive business decisions.



## Figure 2.3 Data view

###  DATA CLEANSING

Data cleansing for the Disney+ Hotstar dataset involves identifying and addressing issues such as missing values, inconsistencies, and inaccuracies to ensure the data is reliable for analysis. Missing values, particularly in key columns like **release\_year**, **genre**, or **age\_rating**, need to be handled by either imputing appropriate values or removing the affected records if the missing data is critical. Inaccurate entries, such as incorrect **release\_years** or implausible **running\_time** values, should be corrected by cross-referencing external sources or removing erroneous records. These steps help improve the integrity of the dataset, making it suitable for analysis and reporting.



## Figure 2.4 Transform data

Additionally, data cleansing includes standardizing attributes such as **age\_rating** and **genre**, ensuring consistency in how values are recorded. For example, variations in age ratings (e.g., "PG-13" vs. "PG 13") and genre names (e.g., "Action/Adventure" vs. "Action Adventure") should be unified to ensure uniformity. Removing duplicate entries is also an essential part of this process, as duplicate records for the same movie or series can distort the analysis. By performing these cleansing steps, the Disney+ Hotstar dataset becomes more accurate, enabling better insights and decision-making regarding content strategies, user engagement, and platform growth.



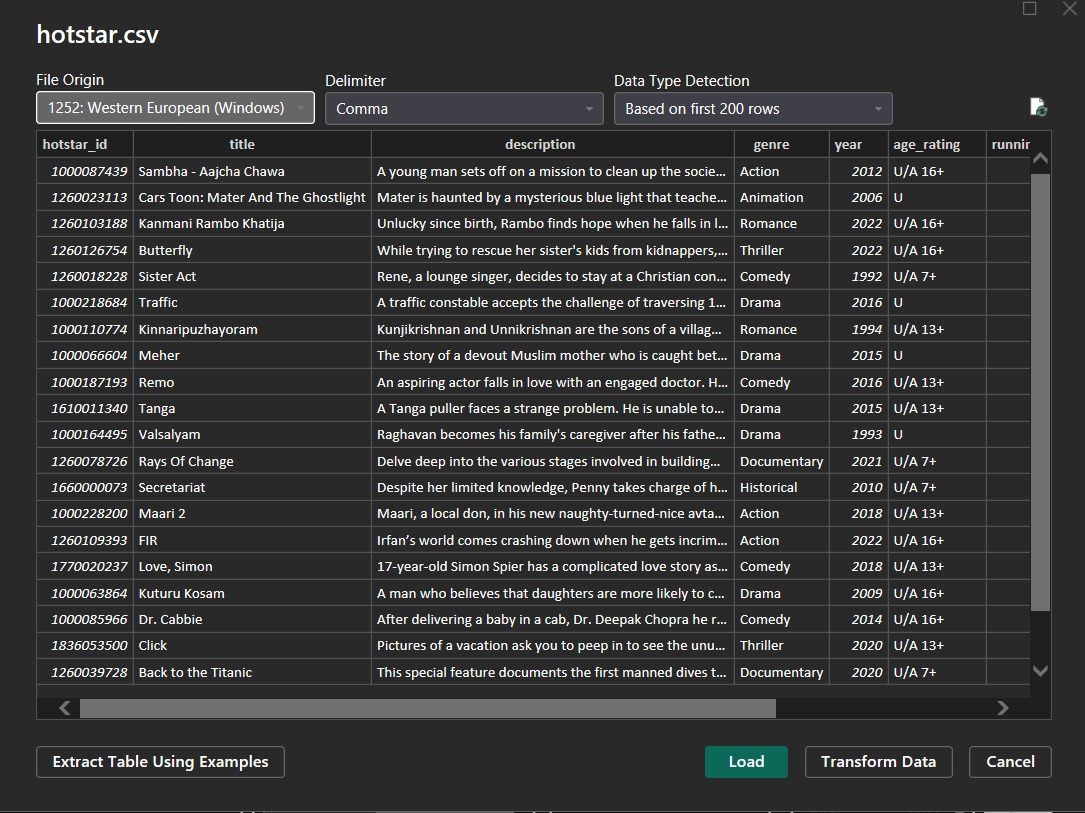
## Figure 2.5 Dax query



## Figure 2.6 Applying the transformations

###  DATA TRANSFORMATION AND ENRICHMENT

Data transformation and enrichment for the Disney+ Hotstar dataset involve converting the data into a more meaningful and analyzable format while also enhancing it with additional relevant information. Transformation processes include converting categorical variables like **genre** or **age\_rating** into numerical or standardized formats, making it easier to perform statistical analysis and aggregations. For instance, genres can be split into multiple columns for each genre category (e.g., Drama, Comedy, Action) using one-hot encoding, allowing for more granular analysis. Additionally, attributes such as **release\_year** can be transformed into time-based features like **decade** or **year\_group** to identify trends over time and observe patterns in content releases across different periods. These transformations help structure the data in a way that makes it suitable for deep analysis and visualization.



## Figure 2.7 Changing datatype

Enrichment involves supplementing the dataset with external information that provides more context or enhances the insights derived from the existing data. For example, integrating data from external sources like IMDb ratings, box office revenue, or viewer ratings could enrich the content’s metadata, allowing for a more comprehensive understanding of content popularity and performance. Enriching the dataset with additional demographic or regional data could also help analyze how different user segments engage with specific genres or content types. Through these transformation and enrichment processes, the Disney+ Hotstar dataset becomes more robust, supporting detailed, insightful analysis of content trends, viewer preferences, and platform performanc

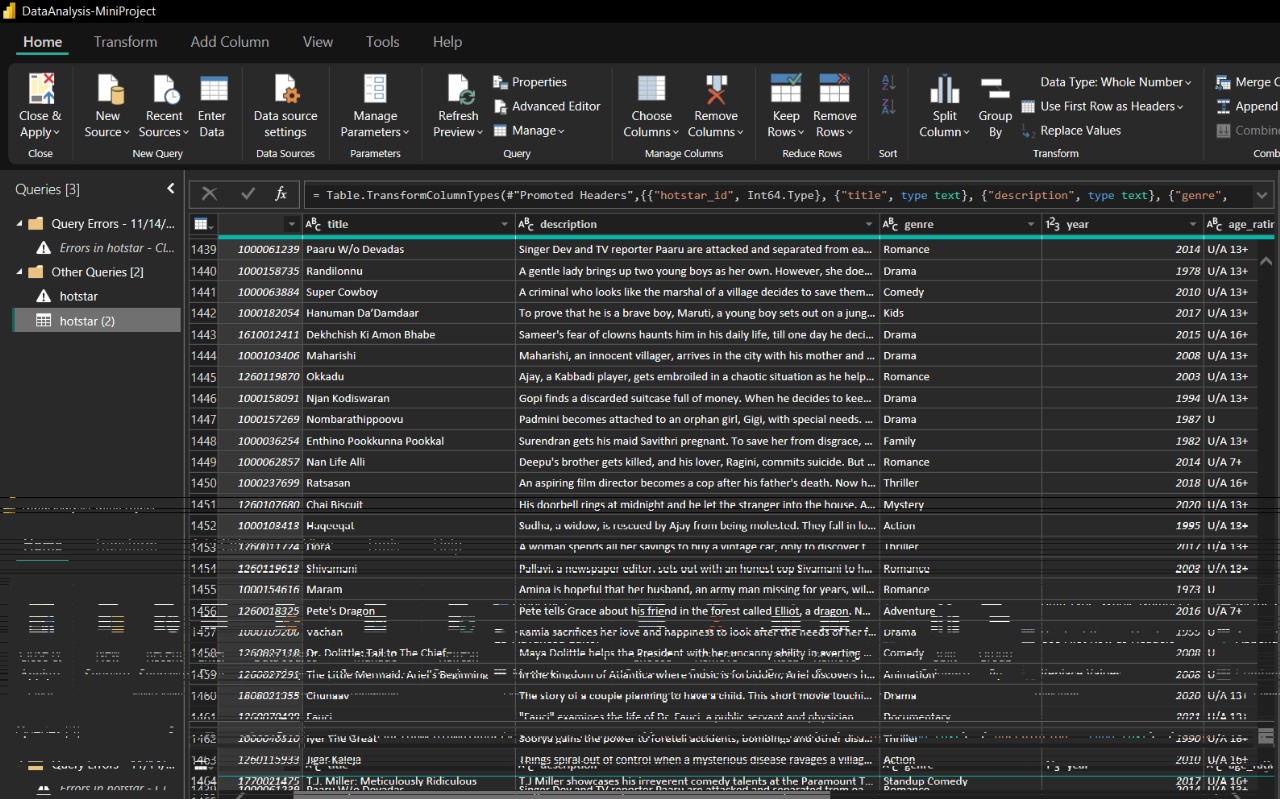


Figure 2.8 Changing name of hostar id

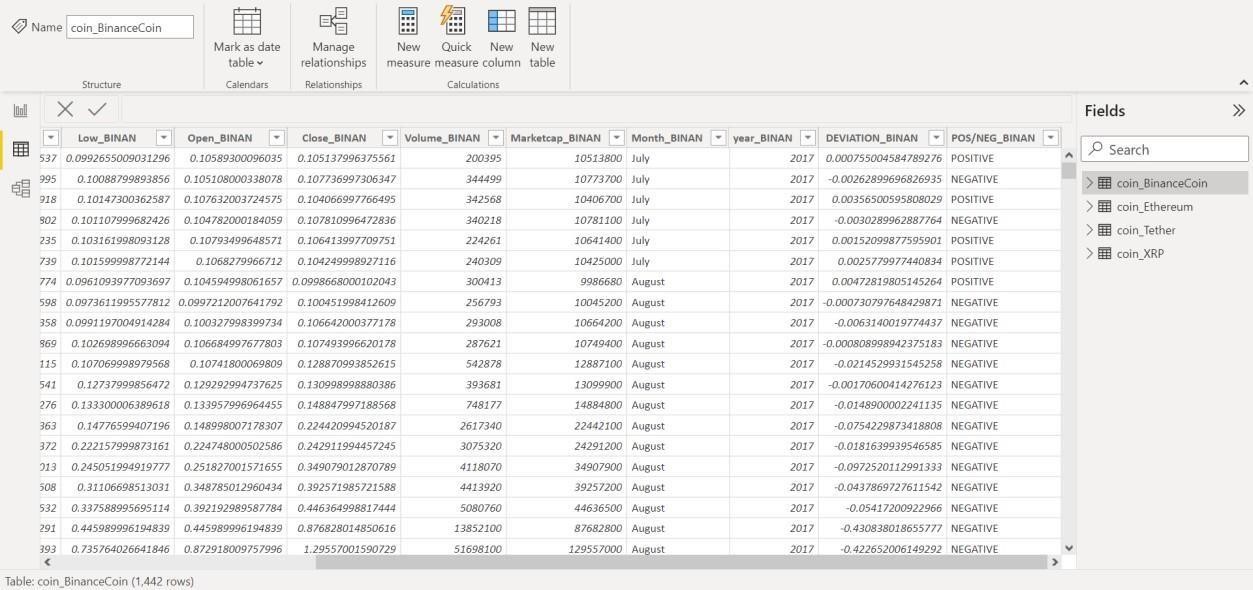


Figure 2.9 Changing name of age\_ratings

###  CHALLENGES ON DATA PREPARATION

Data preparation is a critical step in any analysis, but it comes with several challenges that need careful handling to ensure accurate results. One common issue is **missing or incomplete data**, where fields like running\_time, episodes, or age\_rating may have empty values, making analysis less reliable. This can be addressed by imputing missing values using methods like filling with the mean, median, or mode, or by dropping incomplete rows if they are few and non-critical. Another challenge is **inconsistent data formats**, such as year being stored as text instead of integers or non-numeric entries in the running\_time column. These inconsistencies require conversion to appropriate formats using tools like Power Query or DAX functions. Additionally, **duplicate entries** can inflate counts, so removing duplicates is essential to avoid skewed results.

**Incorrect or outdated values** in fields like genre or age\_rating may also pose problems, necessitating cross-verification with reliable sources. Similarly, **unbalanced categories** can lead to biased insights if some genres dominate the dataset, which can be managed by grouping smaller categories into an “Other” category or applying weighted analysis. Handling **text data** in columns like description can be challenging as it requires preprocessing techniques like tokenization or sentiment analysis for meaningful insights.

Outliers, particularly in running\_time or episodes, can skew averages and distort trends. Detecting and addressing outliers through capping or transformation is important for maintaining data integrity. Moreover, **irrelevant or redundant columns** such as hotstar\_id can clutter the dataset and should be removed if they do not add value. **Data integration issues** may arise when merging datasets due to mismatched entries or incorrect joins, which can be mitigated by using common keys like title or hotstar\_id and validating the merged data. Finally, the dataset may suffer from **granularity mismatches**, such as mixing movies and series, which requires separating analyses based on content type for consistency.

In Power BI, these challenges can be effectively managed using tools like Power Query for cleaning and transforming data, DAX functions for calculated measures, and data validation at each step to ensure accuracy. By addressing these challenges systematically, you can prepare a robust dataset for insightful analysis.

.

# 2.2 DATA MODELLING

Data modeling is the process of organizing and structuring data to support effective analysis and decision-making. In Power BI, it involves creating relationships between tables, defining calculated columns and measures, and establishing hierarchies to simplify data exploration. A well-designed data model ensures that the data is clean, connected, and optimized for performance. Key aspects of data modeling include selecting appropriate primary and foreign keys to establish relationships, defining relationships as one-to-one, one-to-many, or many-to-many, and managing data granularity for consistency. Calculated measures using DAX formulas enable the creation of dynamic metrics like averages, totals, and ratios, while hierarchies (e.g., year, month, day) enhance drill-down capabilities for detailed analysis. Additionally, applying data modeling best practices, such as reducing redundancy through normalization or optimizing query performance with summarization techniques, ensures that the data is both scalable and user-friendly. A robust data model is foundational to creating intuitive and efficient reports in Power BI or any analytical tool.

 TYPES OF DATA MODELS

Conceptual, logical, and physical data models are the three different categories of data models. Conceptual data models, also known as domain models, provide a broad overview of the system's functionality, organizational structure, and business rules. Entity classes, their traits and limitations, their connections, and pertinent security and data integrity requirements are frequently included. Any notation is usually straightforward. Less abstract and offering more information about the ideas and connections in the topic at hand are logical data models.

##  IMPACTS OF DATA MODELS

Need a good data model to acquire the most useful analytics for business intelligence that guides decision making, data modelling and data analytics go hand in hand. Each business unit is forced to consider how they contribute to the overall business goals through the process of developing data models. Additionally, no matter how big and complicated the data estate is or gets, a strong data model ensures improved analytics performance. When all of the data is well-defined, it is considerably simpler to analyze just the data. The linkages between the data attributes have already been established, making it easy to analyze and observe effects as altered the procedures, costs, or staffing.

* DATA MODELING PROCESS

The data modelling process includes identifying the entities of the table, identifying the key properties of each entity, identifying the relationships among entities as the each table contains some relationship among the entities, Mapping the attributes to entities completely, Assigning the keys as needed and deciding on a degree of normalization that balance the needto reduce redundancy with performance requirements and then finalizing and validating the data model .

* BENEFITS OF DATA MODELING

Errors in the creation of databases and software are decreased via data modelling. They substantially improve enterprise wide consistency in documentation and system architecture. Database and application performance are improved through data modelling. They facilitate data mapping across the enterprise and enhance coordination between the business intelligence and development teams. Additionally, they facilitate and expedite the conceptual, logical, and physical phases of database design. The data model relationship of data is defined by default and hence the data model relationship of data set is defined as one-to-one relation for all table as shown in Figure 2.16 .

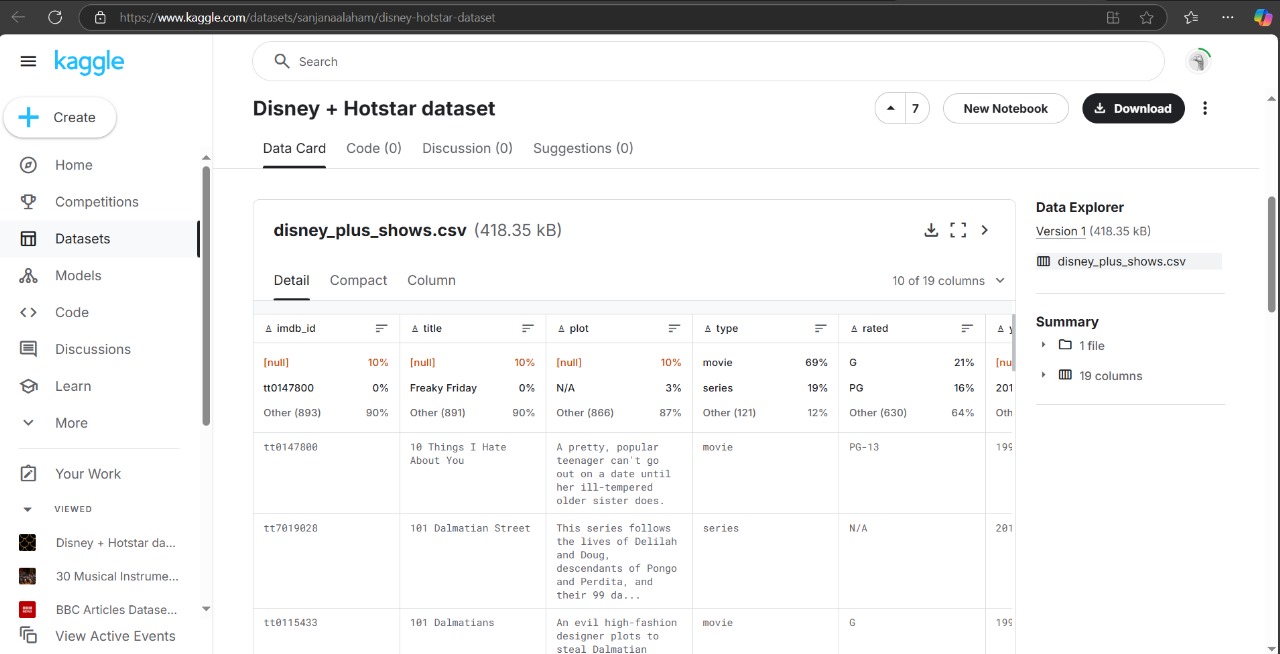


Figure 2.6 Data model

**CHAPTER 3**

**DATA ANALYSIS AND INTERPRETATION**

# 3.1 DATA ANALYSIS

The data analysis process involved a comprehensive exploration of the Disney+ Hotstar dataset to uncover key insights into its content offerings. By analyzing variables such as genre, release year, age rating, and content type, we identified trends in movie and series production over time. This included examining the distribution of genres to determine the most popular categories and analyzing age ratings to understand the target audience. The analysis also revealed patterns in running times, highlighting average durations for different genres and content types. Additionally, by segregating data based on content types like movies and series, we could evaluate the volume and diversity of offerings across these categories.

Further, we investigated season and episode counts for series, providing insights into the production scale and popularity of multi-season content. Using advanced visualizations, we highlighted trends in genre preferences over time and examined correlations between content type and viewer ratings. The results from this analysis will support strategic decisions for enhancing Disney+ Hotstar's content library and tailoring it to meet diverse audience preferences. This data-driven approach helps identify gaps and opportunities, ensuring the platform continues to engage its audience effectively.

.

# QUESTIONS PREPARED FOR THE DATASET

1. Count of Documetary movie&tv show releases from 1928-2022?

1. What is the total count of movies with the age ratings U/A 7+?
2. What is the average running time for the genres StandupComedy, Superhero and Science.

1. What is the AverageEpisode Per Season by each genre and which is the top most?

1. Are there specific genres or types that have a higher frequency of new seasons?
2. What is the total number of seasons in the Animal&Nature Genre?
3. What is the percentage of movies & tv shows available in the Disney+hotstar OTT?

1. What is the name and genre of the movies with longest running time?

1. In which decade there is a sudden rise of the release of the movies?

1. . How many episodes are there between the running time 1-45?

1. In which year the number of movie released is more,( 2020 or 2022),what is the difference between the count in the release of movie for 2020 and 2022?

1. Which genre has the highest average running time in the year 2018?

1. Which movie has the highest running time in the animation genre?

1. Count of tv shows in the year 2019 with the age\_rating "U”?

1. Which is the second highest running time for movies under Adventure Genre in the year 2012?

# 3.2 PUBLISHING DASHBOARDS

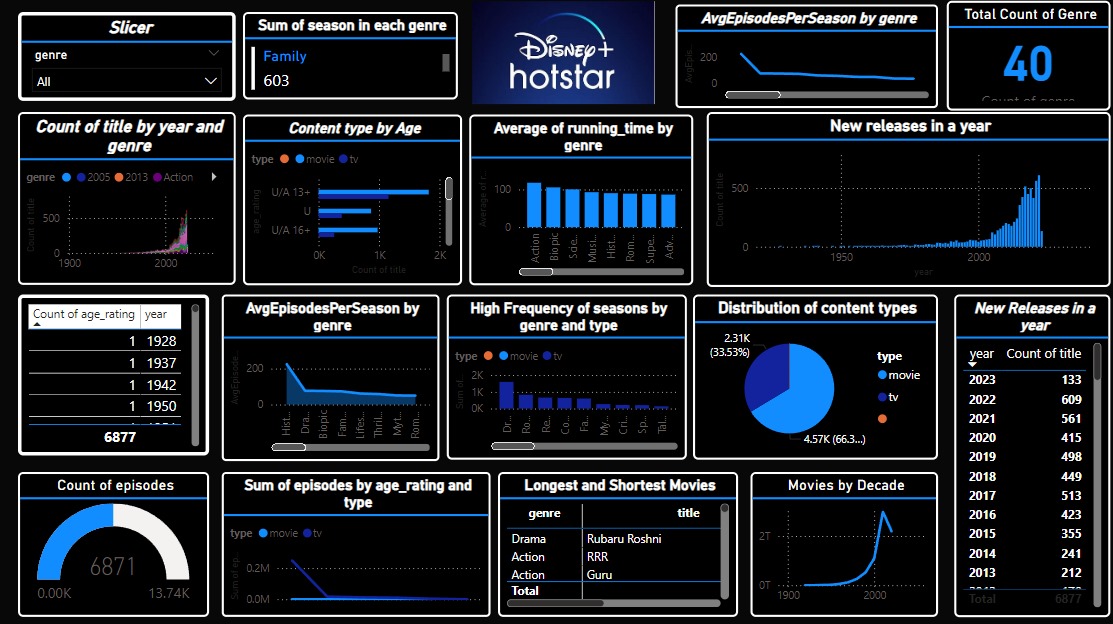
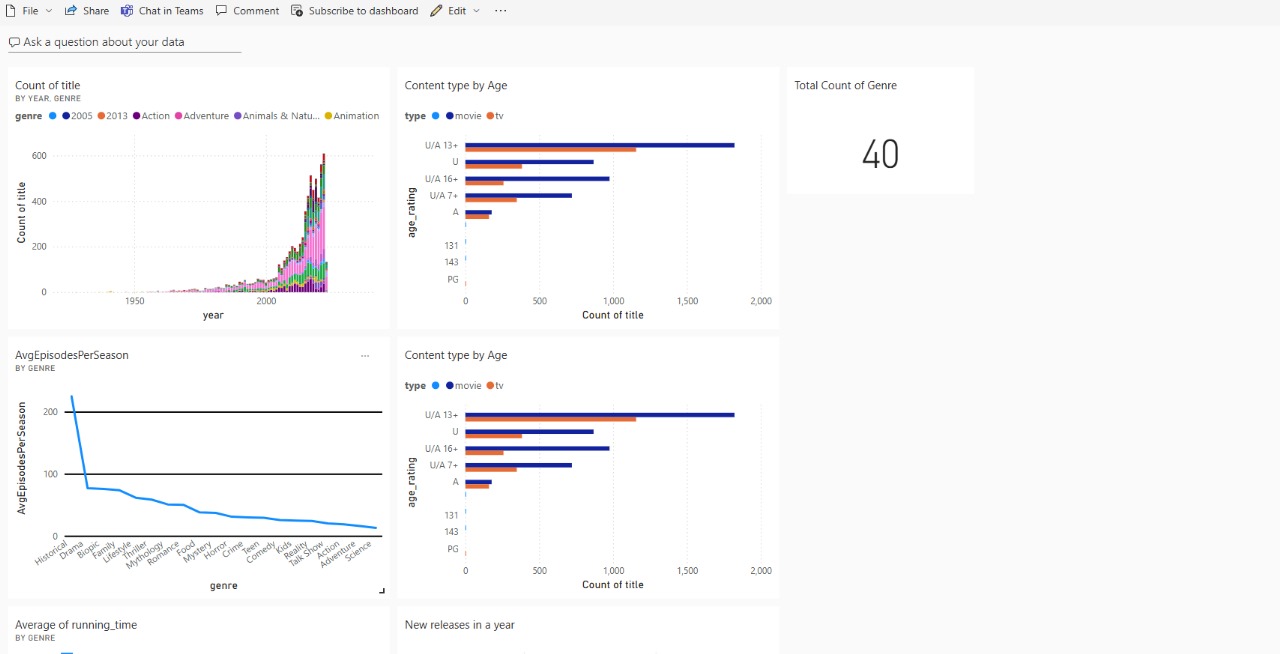
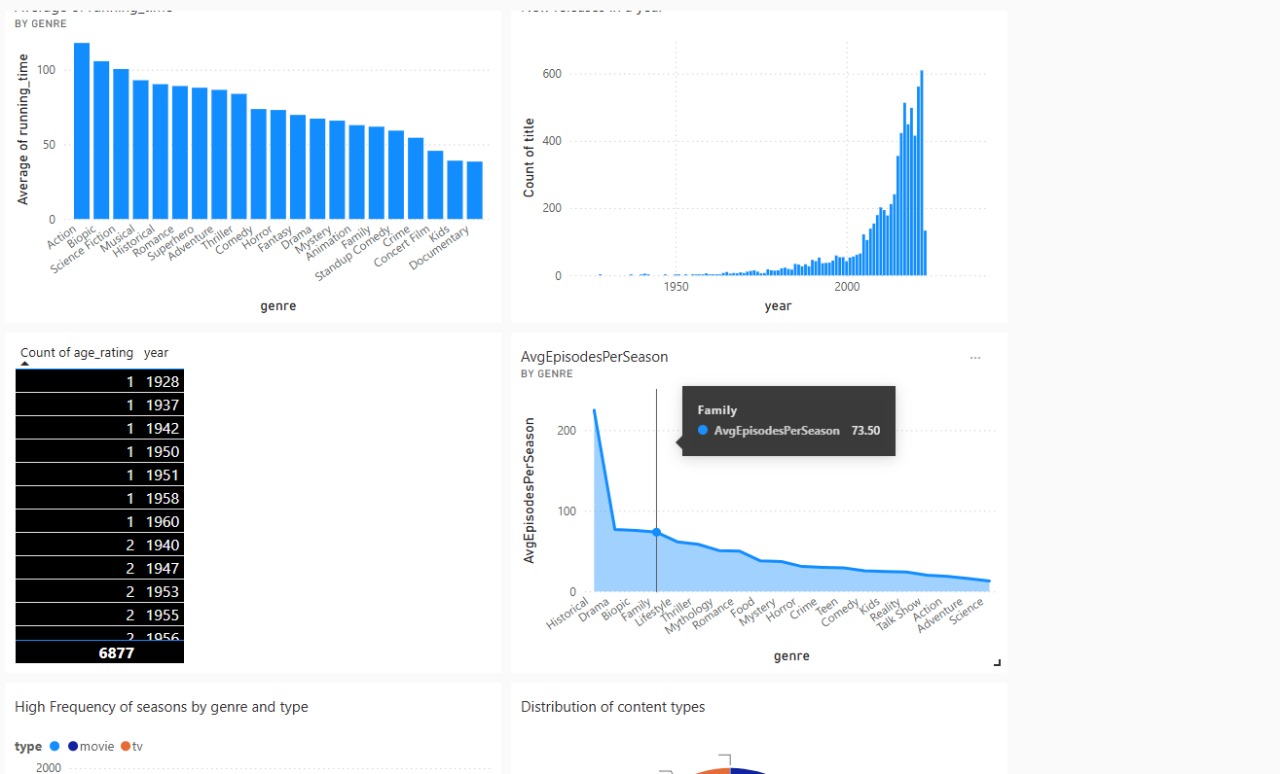


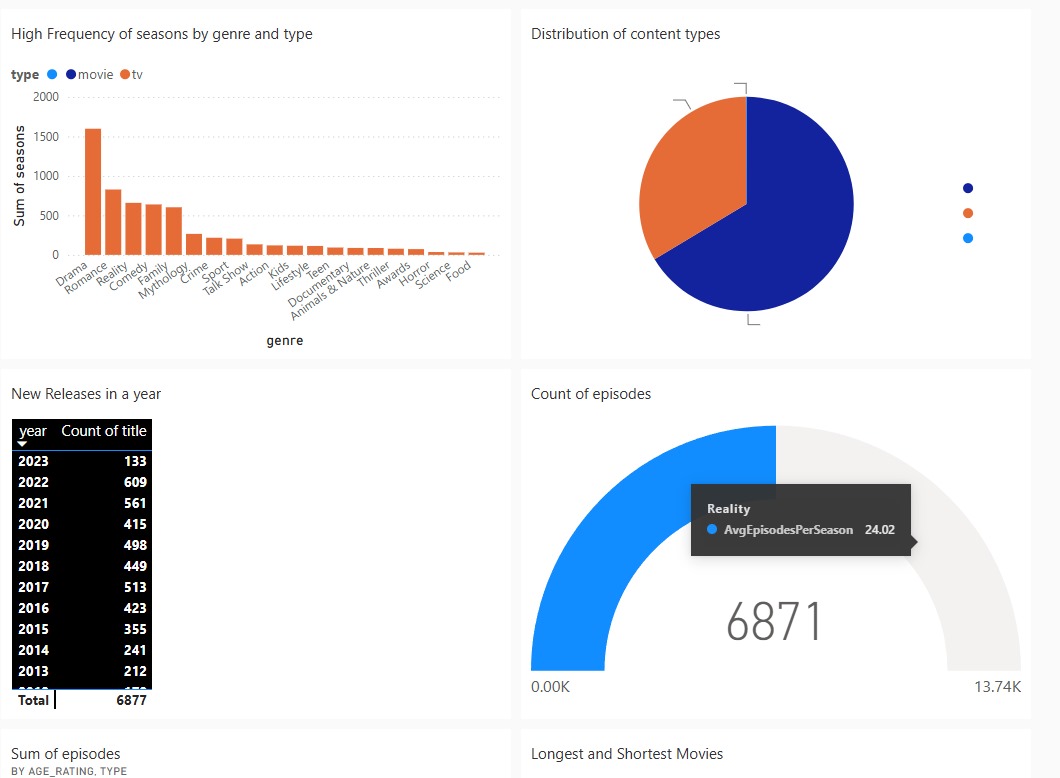
Figure 3.1 Dashboard

**COMPLETE DASHBOARD:**

The Disney+ Hotstar Analysis dashboard is an interactive and insightful visual representation of the platform’s content library, designed using Power BI. This comprehensive dashboard explores various key aspects, such as the average running time across genres, the distribution of content by age ratings, and release trends over the years. It highlights critical insights, such as which genres have the longest average durations, the proportion of movies and series, and the variation in content tailored to different audience age groups. The dashboard also identifies the most popular genres for each year and trends in specific genres over time, providing a clear understanding of Disney+ Hotstar's content strategy. Users can explore the shortest and longest movies, analyze the distribution of genres, and pinpoint the years with the most content releases. By leveraging interactive visualizations like bar charts, pie charts, and line graphs, the dashboard allows users to dive deeper into patterns, trends, and key metrics, offering a detailed view of how Disney+ Hotstar caters to its diverse audience. This analysis serves as a valuable tool for understanding the platform’s strengths and focus areas in its content library.







# 3.3 INFERENCE

1. Count of Documetary movie&tv show releases from 1928-2022?

STEPS:

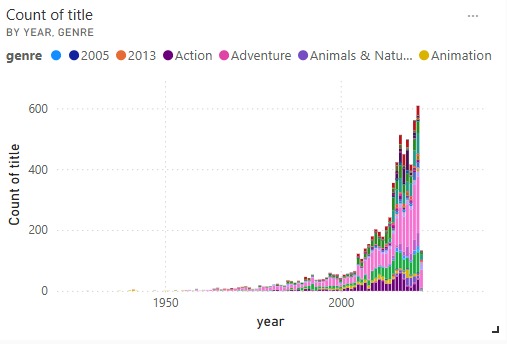
use column chart visualization

x axis - Drag years of the hotstar content

y -axis - Count of Title id (take the id and right click to select count)

slicer - to select the documentary

OUTPUT:



## Figure 3.1 Count of title by year and genre

INFERENCE:

* The year2022 has the maximum count of documentary genre movies&tv shows with the count of 43.

2. What is the total count of movies with the age ratings U/A 7+?

STEPS:

Use Bar Chart

x-axis - Count of titles

y-axis - age\_rating

legend - type

of the dataset

OUTPUT:

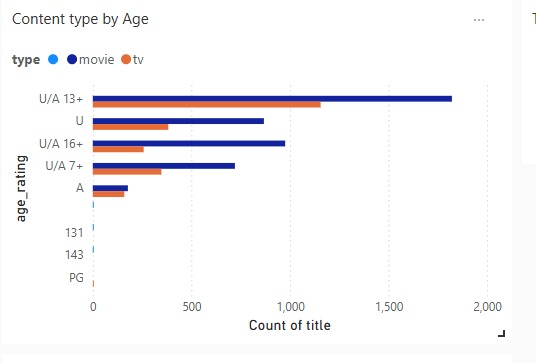


Figure 3.2 Content type by Age

INFERENCE:

Total there are 721 movies with the age rating U/A 7+ between the year 1928-2022

3. What is the average running time for the genres StandupComedy, Superhero and Science.

STEPS:

use Column Chart

x -axis - genre

y-axis - average running time (done by DAX Measure)

OUTPUT:

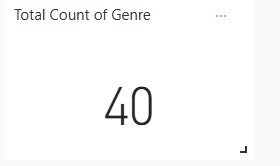


Figure 3.3 Average of running\_time by genre

INFERENCE:

The avg running time for the genre StandupComedy is 59.20, Superhero is 87.95, Science Is 22.56.

4. What is the AverageEpisode Per Season by each genre and which is the top most?

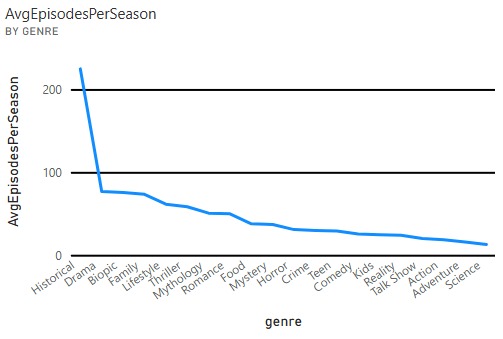
STEPS:

Areachart

x-axis - genre

y-axis - Average Episode Per Season (DAX Measure)

OUTPUT:



## Figure 3.4 AvgEpisodesPerSeason by genre

INFERENCE:

The avgEpdisodePerSeason is achieved using the Stacked area chart and the top most genre is Historical with avg episode 224.

5. Are there specific genres or types that have a higher frequency of new seasons?

STEPS:

Used bar chart

x-axis - genre

y-axis - Sum of Season

legend - type

OUTPUT:

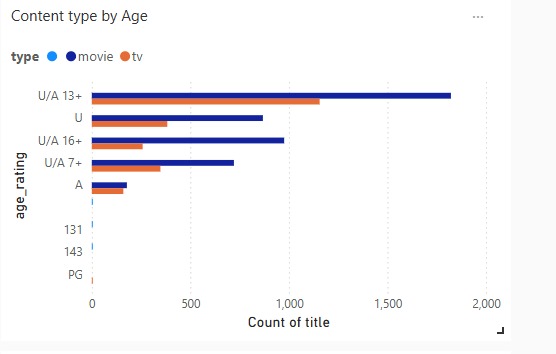


Figure 3.5 High Frequency of seasns by genre and type

INFERENCE:

Drama Genre under tv type has the high frequency of season this done using stacked column chart

6. What is the total number of seasons in the Animal&Nature Genre?

STEPS:

Sum of Seasson in the each genre

using Multi card Visualization

OUTPUT:



## Figure 3.6 Sum of season in each genre

INFERENCE:

* There are nearly 87 season in the Animal&Nature Genre

7. What is the percentage of movies & tv shows available in the Disney+hotstar OTT?

STEPS:

Use Pie chart :

Drag values of type

to get the percentage of type of content in the dataset

OUTPUT:

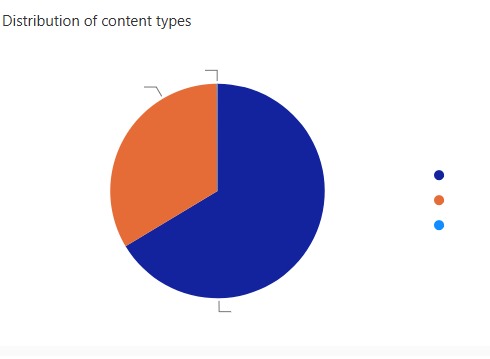


Figure 3.7 Distribution of content types

INFERENCE:

Movies: 66.38% , Tv Shows: 33.35

8. What is the name and genre of the movies with longest running time?

STEPS:

Use Table Visualization

Drag Genre

Drag title

Drag running time for movies

Drag type

to get the Largest and Shortest Movies by running time

OUTPUT:



Figure 3.8 Longest and shortest movies

INFERENCE:

Rubaru Roshni is the movie is the movie with longest running time in the genre Drama

9. In which decade there is a sudden rise of the release of the movies?

STEPS:

use Line chart

x-axis - Decade(DAX Measure)

y axis = Sum of hotstar id

OUTPUT:

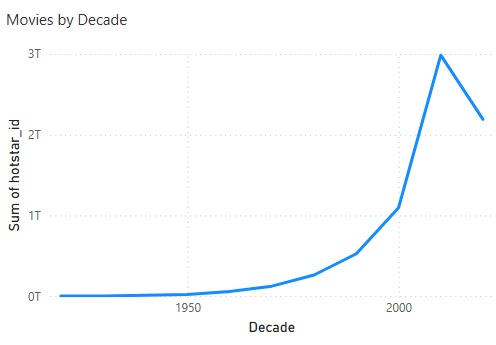


Figure 3.9 movies by Decade

INFERENCE:

Between 2000 to 2010 there is a sudden rise in the release of movies.

10. How many episodes are there between the running time 1-45?

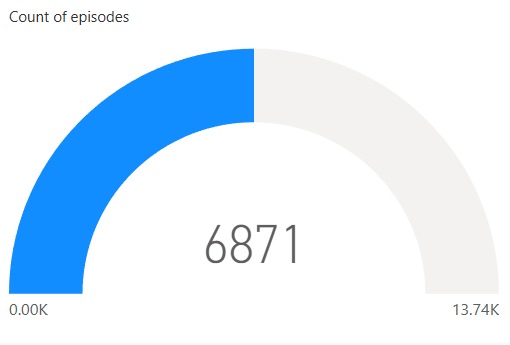
STEPS:

Measure:

Use Guage and Slicer

Drag episodes in gauge and Drag running\_time in slicer

OUTPUT:



## Figure 3.10 Count of episodes

INFERENCE:

There are 1037 episodes between the running time 1-45.

1. In which year the number of movie released is more,( 2020 or 2022),what is the difference between the count in the release of movie for 2020 and 2022?

STEPS:

Use Table

Drag year

Drag Count of title

OUTPUT:

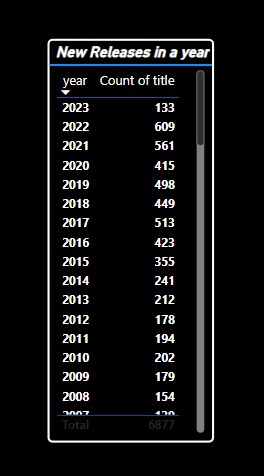


Figure 3.11 New releases on a year

INFERENCE:

In the 2020 415 movies where released ,in 2022 609 movies were released , In the year 2022 194 movies were released more than the year 2020.

1. Which genre has the highest average running time in the year 2018?

STEPS:

Use Table

Drag year

Drag Count of title

for bar chart

x-axis - genre

y-axis - Average Running Episodes

OUTPUT:

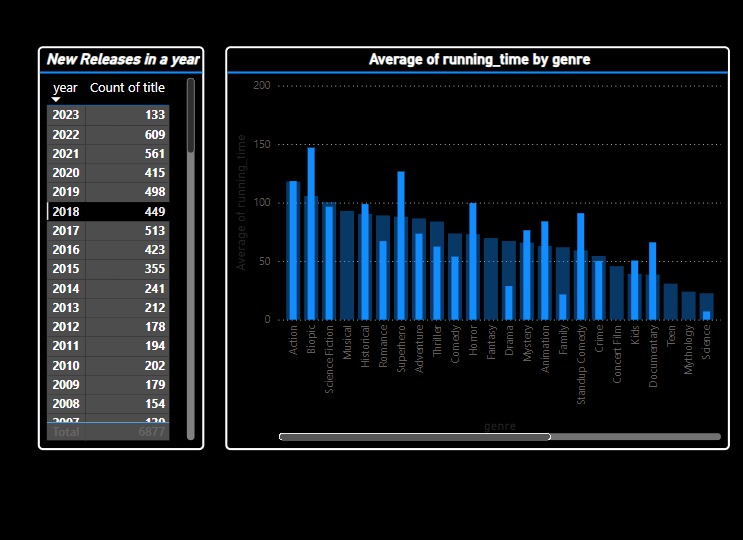


Fig 3.12 Average of running\_time by genre

INFERENCE:

Biopic Genre has the highest running time in the year 2018.

13. Which movie has the highest running time in the animation genre?

STEPS:

Use Slicer

Drag Genre

Use Table Visualization

Drag Genre

Drag title

Drag running time for movies

Drag type

to get the Largest and Shortest Movies by running time

OUTPUT:



## Figure 3.13 Longest and shortest movies by slicer

INFERENCE:

Bal Ganesh has the highest running time in the Animation Genre.

14. Count of tv shows in the year 2019 with the age\_rating "U"?

STEPS:

Use Table

Drag year

Drag Count of title

Use Bar Chart

x-axis - Count of titles

y-axis - age\_rating

legend - type

of the dataset

OUTPUT:

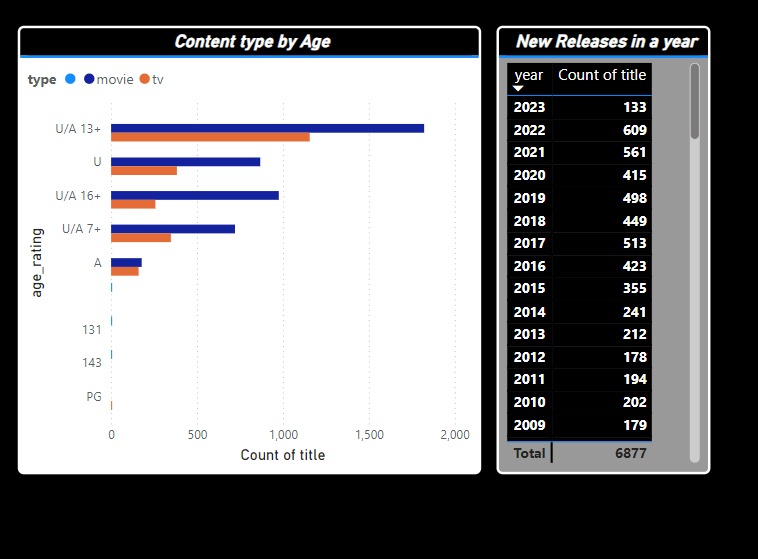


Figure 3.14 Content type by age release year

INFERENCE:

* + Count of tv shows in the year 2019 with the age\_rating "U" is 70It increased during the year 2016 to 2018

15. Which is the second highest running time for movies under Adventure Genre in the year 2012?

STEPS:

Use Table Visualization

Drag Genre

Drag title

Drag running time for movies

Drag type

to get the Largest and Shortest Movies by running time

Use Table

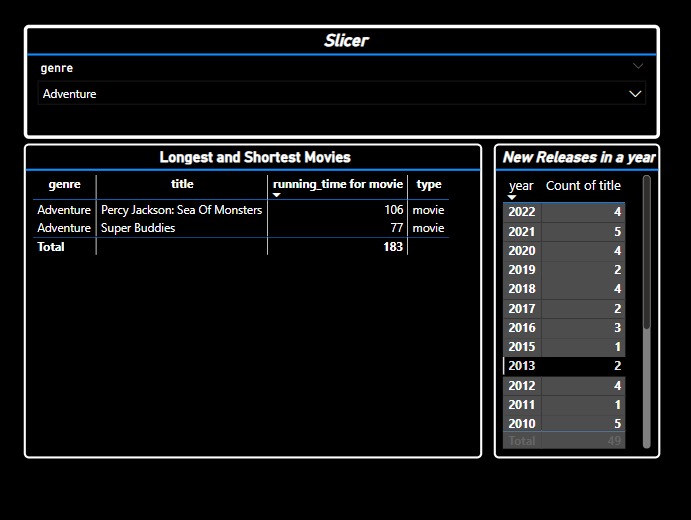
Drag year

Drag Count of title

Use Slicer

Drag Genre

OUTPUT:



# Figure 3.15 Longest and shortest movies slicer and release year

INFERENCE:

Super Buddies is the movies wuth second highest running time for movies under Adventure Genre in the year 2012.

**CHAPTER 4**

**CONCLUSION**

In conclusion, **data modeling** is an indispensable process in data analytics that bridges the gap between raw data and actionable insights. It serves as the foundation for organizing and structuring data, enabling seamless integration, efficient querying, and accurate analysis. By defining relationships, establishing hierarchies, and leveraging calculated fields or measures, data modeling ensures that datasets are not only comprehensible but also aligned with business objectives. Properly designed models, such as those following star or snowflake schema principles, optimize performance, reduce redundancy, and make reports intuitive for end-users.

Furthermore, the scalability of a well-constructed data model allows it to adapt to evolving data complexities without compromising accuracy or efficiency. In tools like Power BI, data modeling empowers users to explore their datasets with ease, providing a robust framework for advanced analytics and insightful visualizations. It also ensures that the data's integrity is maintained, even when dealing with complex relationships or large datasets.

Ultimately, data modeling transforms disparate data sources into a coherent structure, enabling organizations to derive meaningful insights that drive informed decision-making. As the backbone of any data-driven initiative, investing time and effort into building a strong data model pays off by ensuring accurate, efficient, and scalable analytics for long-term success.

**REFERENCES**

1. Choudhary, Shailee, Dr Rinku Sharma Dixit, and Nikhil Govil. "Analysing Consumer Preferences and Pattern for Ott Platforms: A Study." Available at SSRN 4605161 (2023).

1. [Chomski, Joseph. "Case study: Power bi visualizations applied to digital video publisher's adveryising space." Computer Science and Education in Computer Science 14, no. 1 (2018): 139-145.](https://www.investopedia.com/terms/c/cryptocurrency.asp)

1. Bhargava, Mandava Geetha, K. T. P. S. Kiran, and Duvvada Rajeswara Rao. "Analysis and design of visualization of educational institution database using power bi tool." Global Journal of Computer Science and Technology 18, no. C4 (2018): 1-8.[.](https://www.investopedia.com/analyze-crypto-6456223#%3A~%3Atext%3DA%20cryptocurrency%20analysis%20involves%20examining%2Cto%20determine%20its%20trends%20better)
2. Anand, Akshat, D. Sarika, Harshit Malviya, Dipti Vatsa, Sunita Adarsh Yadwad, and Shreya M. Dholariya. "Churn Buster: Uncovering Patterns and Predicting Churn in OTT Platforms." In 2024 Parul International Conference on Engineering and Technology (PICET), pp. 1-6. IEEE, 2024.
3. Gonçalves, Célia Talma, Maria José Angélico Gonçalves, and Maria Inês Campante. "Developing Integrated Performance Dashboards Visualisations Using Power BI as a Platform." Information 14, no. 11 (2023): 614.
4. Stefanova, Tanja, Zoran Zdravev, and Aleksandar Velinov. "Analysis of Top Selling Products Using Business Intelligence." Balkan Journal of Applied Mathematics and Informatics 6, no. 2 (2023): 79-90.