



Transform Yourself

# **DISNEY HOTSTAR ANALYSIS**

## **PROJECT GUIDE**

Dr.P.Natesan Professor,AI

## **PROJECT MEMBER**

K.Neeleshwar-22ADR074

K.Rahul-22ADR084

P.Ranjith Kumar-22ADR086



# OBJECTIVES

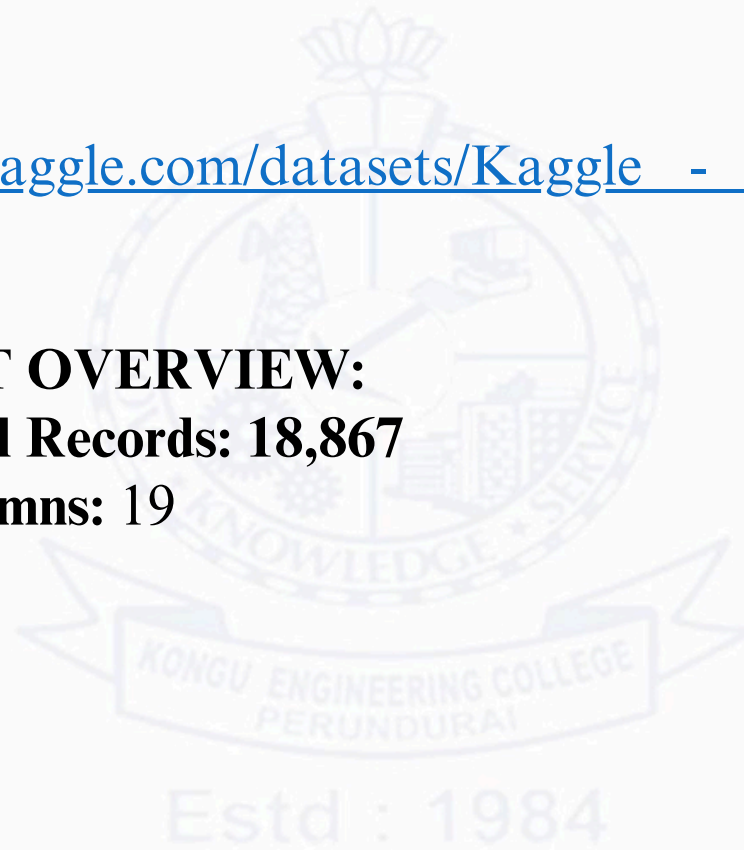
- The analysis of Disney Plus's video library offers a detailed exploration of key metrics such as release year, genre, runtime, and IMDb ratings. By visualizing trends in popular genres, production timelines, and content types (movies vs. series), we can uncover insights into audience engagement and preferences.
- Examining the distribution of genres, ratings, and release patterns provides a deeper understanding of viewer demand and the platform's strategic focus. Furthermore, analyzing Disney Plus's growth over time and genre performance identifies areas of strength and potential gaps, enabling data-driven recommendations for expanding and diversifying the content library to enhance appeal and maintain market competitiveness.

# DATASET LINK

- <https://www.kaggle.com/datasets/Kaggle - Disney+ Hotstar Dataset>

- **DATASET OVERVIEW:**

- **Total Records: 18,867**
- **Columns: 19**



# DATASET DESCRIPTION

- Type of Content: The dataset distinguishes between movies and series, offering insights into their relative proportions.
- Genres: A variety of genres are represented, enabling analysis of popular categories and their performance.
- Temporal Information: The release year and specific release dates allow analysis of trends over time.
- Creators and Cast: Includes details about directors and main cast members, helping analyze the influence of key contributors on content success.
- This dataset is optimized for exploring Disney Plus's content offerings, focusing on production trends, genres, and runtime characteristics.

# DATA PREPROCESSING

- Handle Missing Values
- Replace or remove missing data based on relevance (e.g., drop rows with missing title, fill director with "Unknown").
- Ensure Correct Data Types
- Convert release\_year to integer and released\_at to datetime.
- Ensure runtime is numeric.
- Standardize Text
- Convert text columns (title, genre, director, cast) to lowercase and trim whitespace.
- Split and Extract Data
- Extract month and year from released\_at.
- Split multi-genre entries into individual rows.
- Address Outliers
- Fix extreme or invalid values in runtime and release\_year.
- Remove Duplicates
- Eliminate duplicate rows to maintain data integrity.



# QUESTIONS FOR DATA ANALYSIS

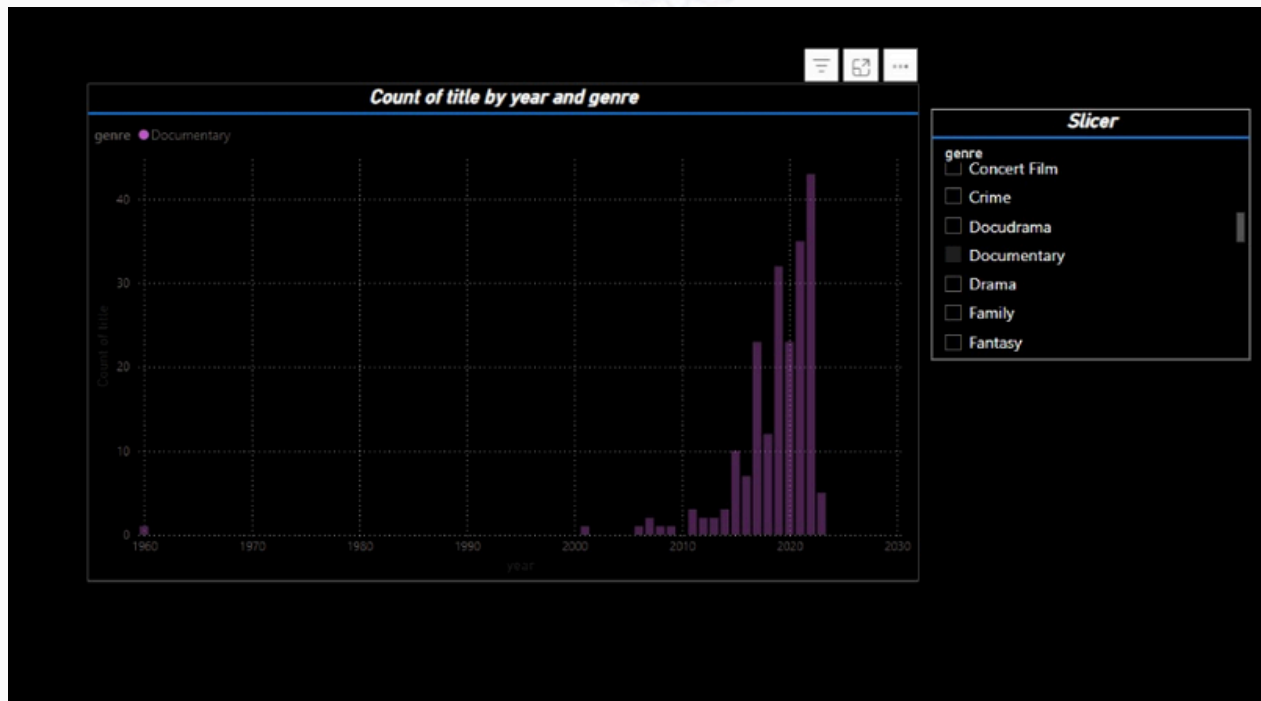
1. Count of Documetary movie&tv show releases from 1928-2022?
2. What is the total count of movies with the age ratings U/A 7+?
3. What is the average running time for the genres StandupComedy, Superhero and Science.
4. What is the AverageEpisode Per Season by each genre and which is the top most?
5. Are there specific genres or types that have a higher frequency of new seasons?
6. What is the total number of seasons in the Animal&Nature Genre?
7. What is the percentage of movies & tv shows available in the Disney+hotstar OTT?
8. What is the name and genre of the movies with longest running time?

# QUESTIONS FOR DATA ANALYSIS

1. In which decade there is a sudden rise of the release of the movies?
2. How many episodes are there between the running time 1-45?
3. 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?
4. Which genre has the highest average running time in the year 2018?
5. Which movie has the highest running time in the animation genre?
6. Count of tv shows in the year 2019 with the age\_rating "U"?
7. Which is the second highest running time for movies under Adventure Genre in the year 2012?

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

Visualization:

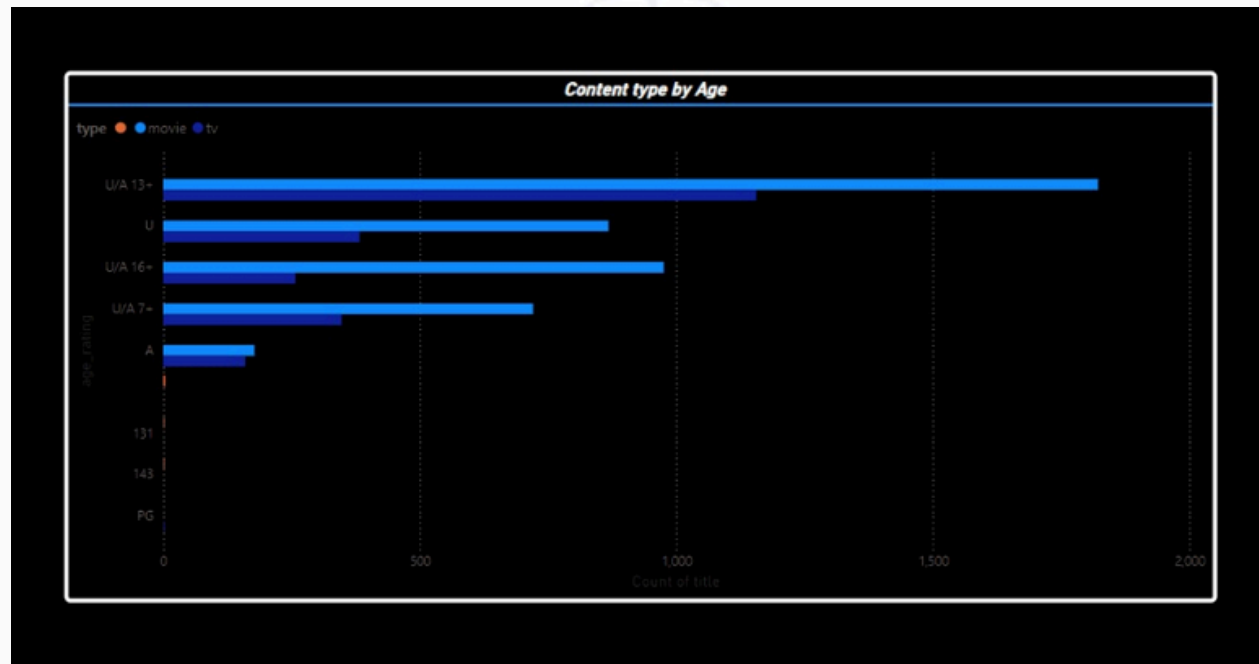


*Inference:* The year 2022 has the maximum count of documentary genre movies&tv shows with the count of 43.



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

Visualization:

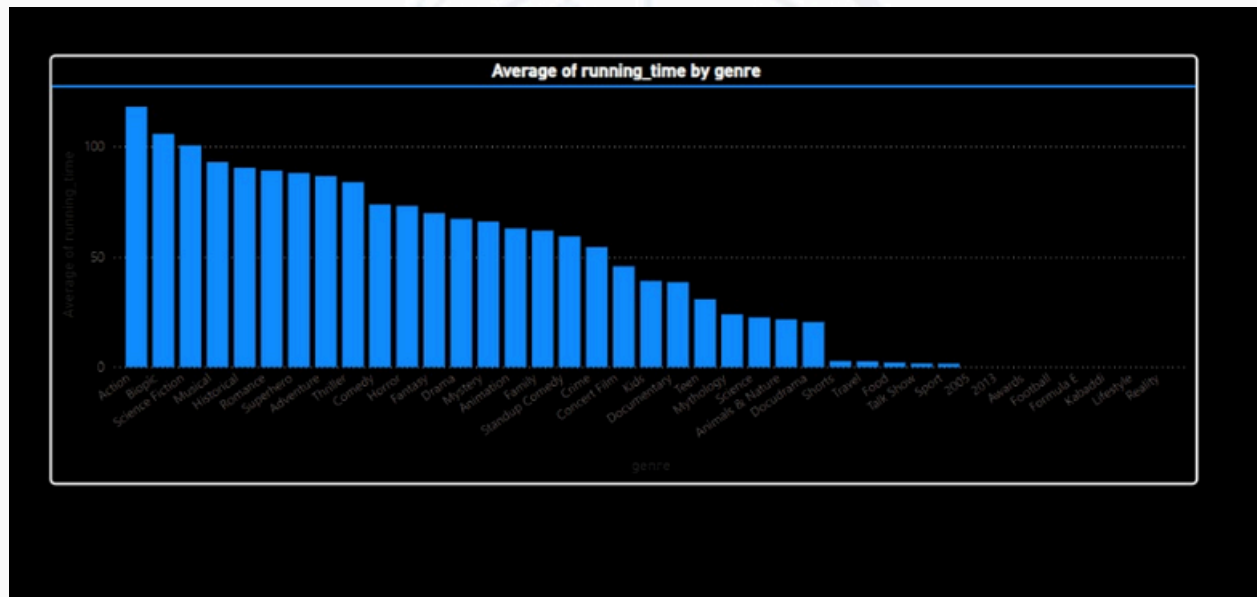


Estd : 1984

*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.

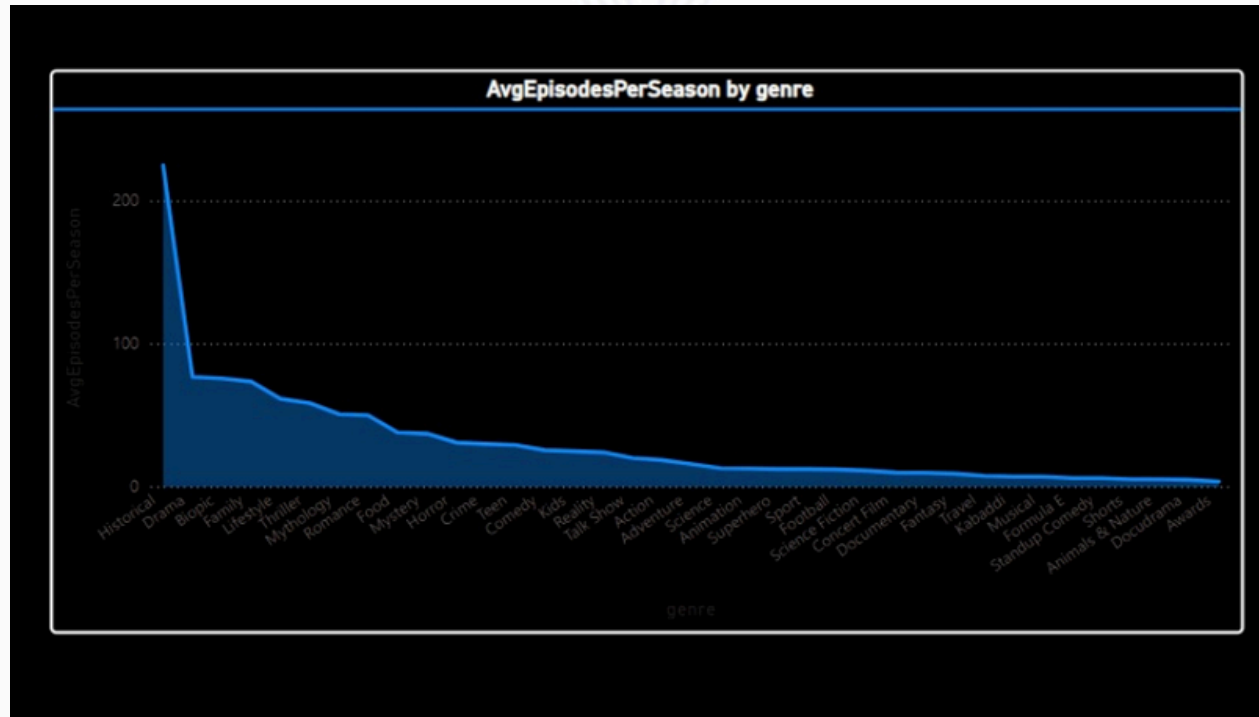
Visualization:



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?**

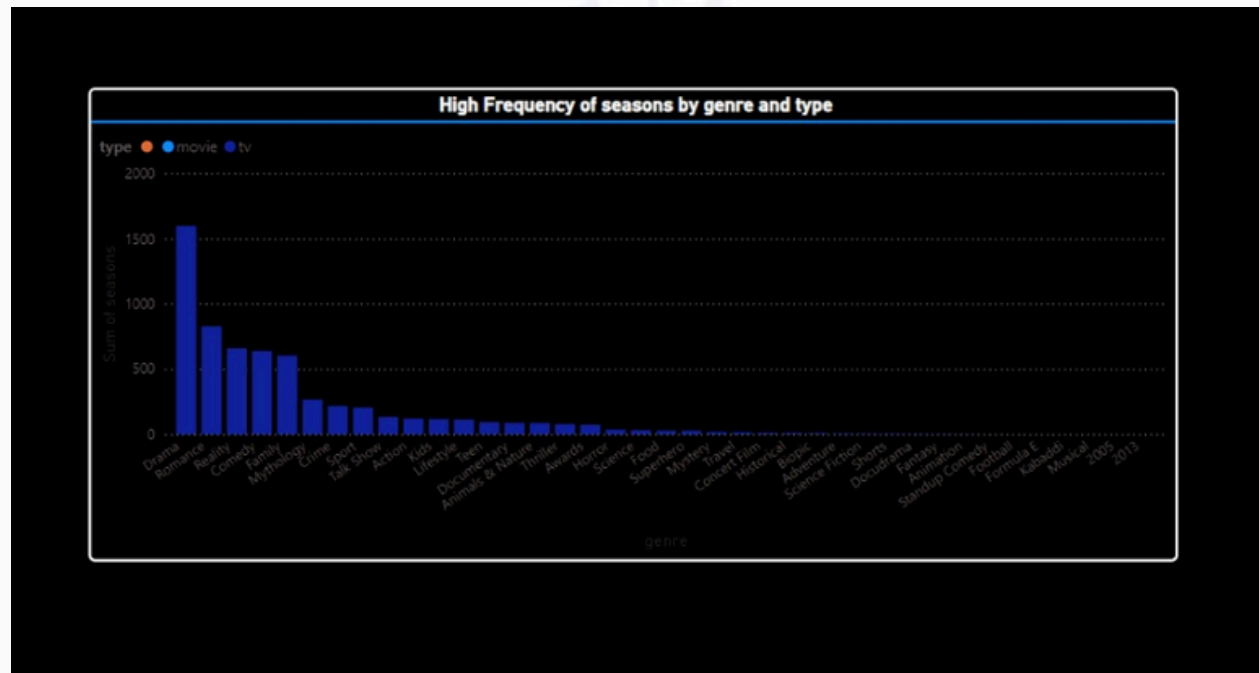
Visualization:



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?**

Visualization:



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?

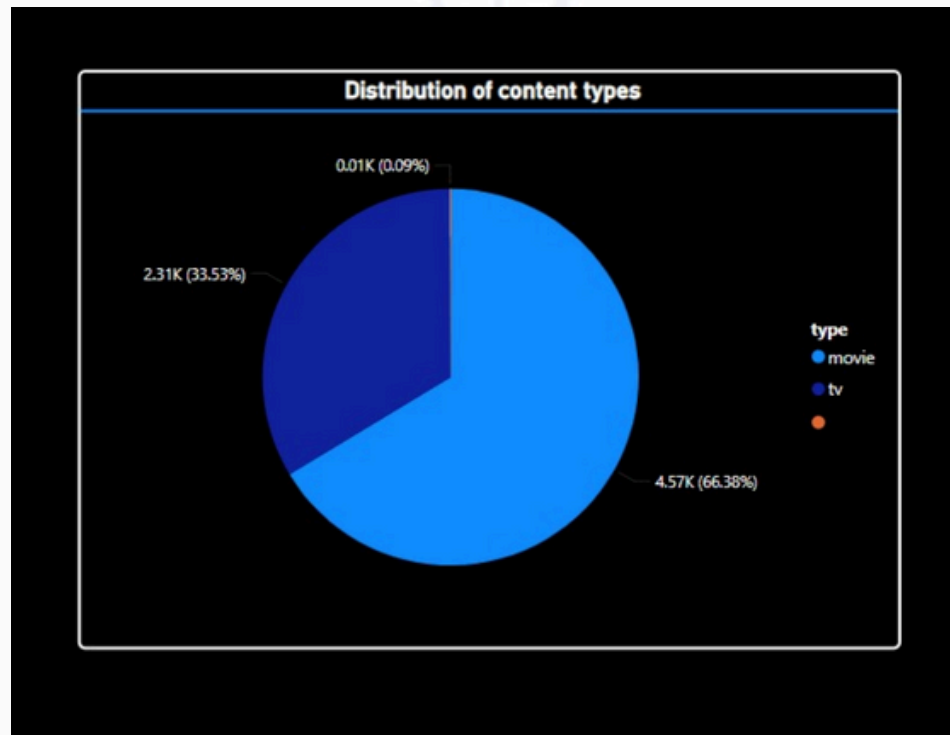
Visualization:



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?**

Visualization:




Inference : Movies: 66.38% , Tv Shows: 33.35



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

Visualization:

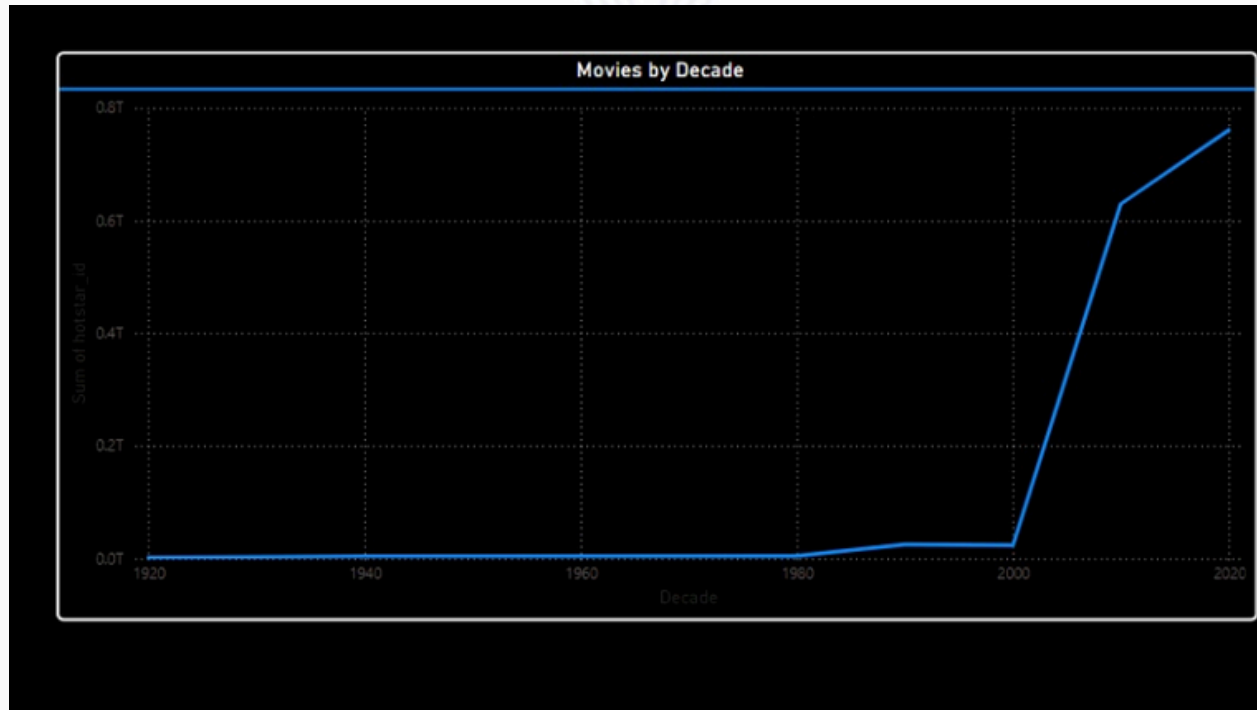


genre	title	running_time for movie	type
Drama	Rubaru Roshni	770	movie
Action	RRR	709	movie
Action	Guru	567	movie
Biopic	MS Dhoni: The Untold Story	532	movie
Romance	Kanmani Rambo Khatija	471	movie
Action	Vishwaroopam 2	404	movie
Action	Baahubali 2: The Conclusion	320	movie
Thriller	Drishyam	314	movie
Romance	Raja Rani	314	movie
Romance	Raja Rani	314	tv
Romance	Satya In Love	300	movie
Action	Saamy 2	296	movie
Action	Tony	296	movie
Action	Bang Bang	292	movie
Romance	Premam	290	movie
Action	Vikrant Rona	290	movie
Drama	Bandhan	289	movie
Action	Pulimurugan	289	movie
Drama	Chandelakha	288	movie
Total		450679	

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?

Visualization:



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?**

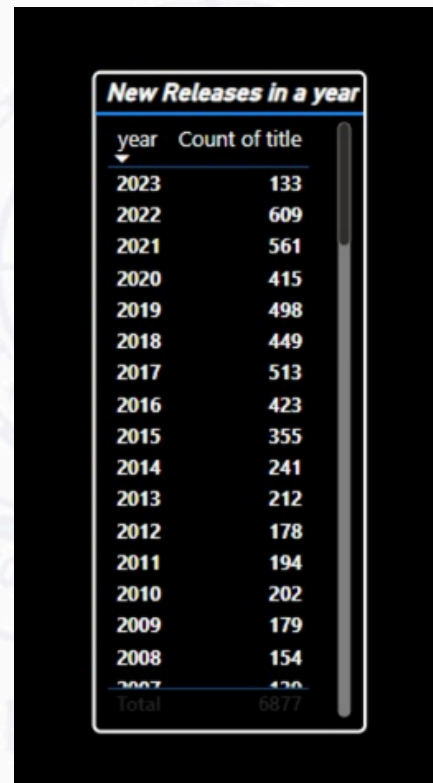
Visualization:



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

11) 11. 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?

Visualization:

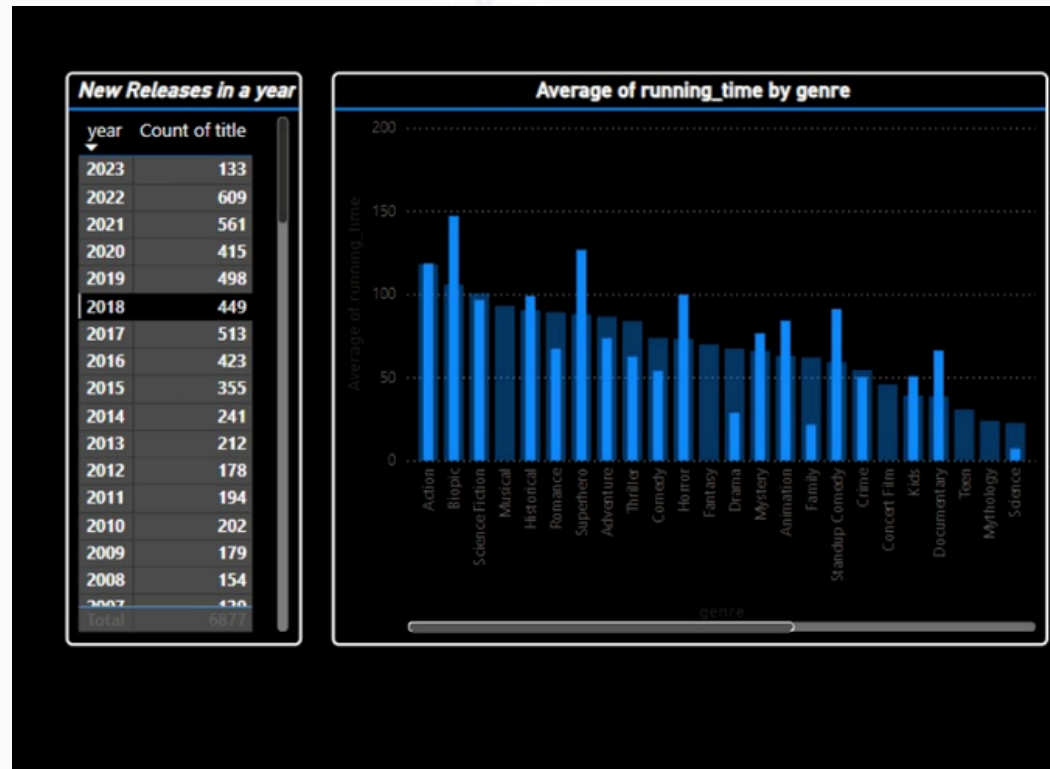


year	Count of title
2023	133
2022	609
2021	561
2020	415
2019	498
2018	449
2017	513
2016	423
2015	355
2014	241
2013	212
2012	178
2011	194
2010	202
2009	179
2008	154
2007	130
Total	6877

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.

12) Which genre has the highest average running time in the year 2018?

Visualization:



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

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

Visualization:

*Slicer*

genre  
Animation

**Longest and Shortest Movies**

genre	title	running_time for movie	type
Animation	Bal Ganesh	201	movie
Animation	Dory's Reef Cam	181	movie
Animation	Arendelle Castle Yule Log	180	movie
Animation	Arendelle Castle Yule Log: Cut Paper Edition	180	movie
Animation	Bal Ganesh and Friends from Zeba	156	movie
Animation	Incredibles 2	119	movie
Animation	My Friend Ganesha	118	movie
Animation	Cars	116	movie
Animation	The Incredibles	115	movie
Animation	Fantasia	114	movie
Animation	Ralph Breaks The Internet	113	movie
Animation	Ratatouille	111	movie
Animation	Moana	107	movie
Animation	Raya and the Last Dragon	107	movie
Animation	Cars 2	106	movie
Animation	Coco	105	movie
Animation	Frozen II	104	movie
Total		15085	

Inference : The average runtime for TV shows is 45 minutes.



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

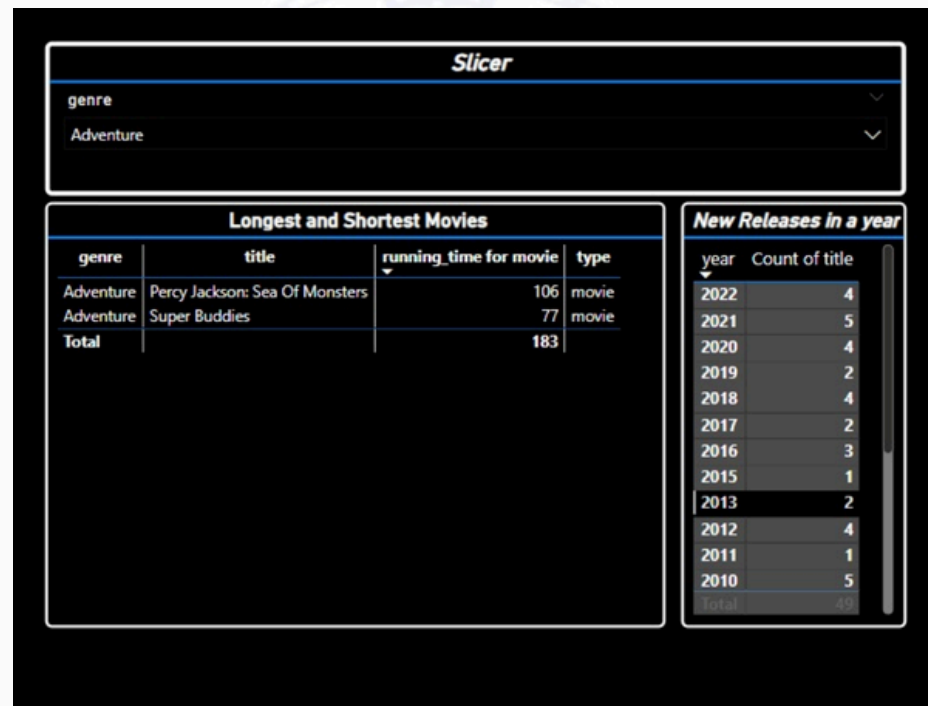
Visualization:



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

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

Visualization:



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

**THANK  
YOU**

