BIG DATA AND BUSINESS INTELLIGENCE MODULE

TOPIC- NETFLIX MOVIES AND TV SHOWS ANALYSIS

SECTION 1: BUSINESS INTELLIGENCE DESIGN

TECHNICAL REPORT FOR DEVELOPERS

NAME: **RAJASHEKAR REDDY KORIMILLA**

Student ID- **W9633347**

Submission Date- **10-01-2024**

Table of contents

Section 1: Business Intelligence Design

Section 2: Business Intelligence Solution

Section 1: Business Intelligence Design

Introduction

Dataset description

Overview of business questions

BI Data Pre-processing

    Data loading

    Data cleaning

Custom columns with power query editor

Modelling and creating relationships

Custom columns with DAX calculation's

Data source description and bi requirements

Introduction

Netflix is the world's largest online streaming platform with over 220 million subscribers in over 190 countries enjoying movies tv series, documentaries, feature films across a wide variety of genres and languages. around the world as of 2022. it important to see how effectively they merge the movies and tv shows on their platform in order to enhance the user experience, without losing subscribers churn.

Data set source and description

The datasets used is the Netflix movies and tv shows which obtained from Kaggle https://www.kaggle.com/datasets/shivamb/netflix-shows

https://www.kaggle.com/datasets/pariaagharabi/netflix2020

https://www.kaggle.com/code/arnavsmayan/netflix-userbase-visualization-notebook/input

This dataset contains movies and tv shows from 1969-2022 including title release year, age, certification, runtime, genres, production countries, seasons information for tv shows, IMDB score, number of votes, Netflix userbase, total revenue and subscribers.

Screenshot of the datasets are presented below:

Best Shows Netflix

index	TITLE	RELEASE_	SCORE	NUMBER_	DURATION	NUMBER_	MAIN_	GET	MAIN_PRODUCTION
244	All of Us A	2023	7.5	41393	61	1	action	KR	
223	Outer Ban	2020	7.5	43404	49	2	action	US	
229	Ragnarok	2020	7.5	36185	47	2	action	NO	
224	Arrow	2012	7.5	425716	42	8	action	US	
200	Teenage B	2020	7.6	10821	49	1	action	US	
197	Alice in Bo	2020	7.6	47651	47	2	action	JP	
182	Midnight h	2021	7.7	102321	64	1	action	US	
159	A Series of	2017	7.8	59239	47	3	action	US	
172	NCIS	2003	7.8	141049	44	10	action	US	
138	Kill la Kill	2013	7.9	13372	25	1	action	JP	
140	Into the Ba	2015	7.9	45618	43	3	action	US	
136	The Seven	2014	7.9	30341	24	5	action	JP	
141	Merlin	2008	7.9	80138	44	5	action	GB	
142	InuYasha	2000	7.9	25823	25	5	action	JP	
126	Squid Gam	2021	8	459718	54	2	action	KR	
129	Marco Pol	2014	8	71239	55	2	action	US	
102	The Walkin	2010	8.2	945125	46	11	action	US	
75	Kingdom	2019	8.3	43760	48	2	action	KR	
66	Vincenzo	2021	8.4	55134	81	1	action	KR	
58	Ash vs Evil	2015	8.4	70087	30	3	action	US	
62	Trollhunte	2016	8.4	16509	22	3	action	US	
64	The Legen	2012	8.4	117464	23	4	action	US	
44	The Last K	2015	8.5	129479	55	5	action	GB	
40	Sacred Gar	2018	8.6	85088	50	2	action	IN	
35	Cobra Kai	2018	8.6	164858	31	5	action	US	
25	One Piece	2015	8.7	148388	24	2	action	JP	
18	One Piece	1999	8.8	312586	24	21	action	JP	
3	Arcane	2021	9.1	175412	41	1	action	US	

Netflix Userbase

User ID	Subscripti	Monthly R	Join Date	Last Payment Date	Country	Age	Gender	Device	Plan Duration
1	Basic	10	15-01-2022	10-06-2023	United Sta	28	Male	Smartphone	1 Month
2	Premium	15	05-09-2021	22-06-2023	Canada	35	Female	Tablet	1 Month
3	Standard	12	28-02-2023	27-06-2023	United Kin	42	Male	Smart TV	1 Month
4	Standard	12	10-07-2022	26-06-2023	Australia	51	Female	Laptop	1 Month
5	Basic	10	01-05-2023	28-06-2023	Germany	33	Male	Smartphone	1 Month
6	Premium	15	18-01-2023	27-06-2023	France	25	Female	Smart TV	1 Month
7	Standard	12	09-12-2021	25-06-2023	Brazil	46	Male	Tablet	1 Month
8	Basic	10	02-04-2023	24-06-2023	Mexico	39	Female	Laptop	1 Month
9	Standard	12	20-10-2022	23-06-2023	Spain	37	Male	Smartphone	1 Month
10	Premium	15	07-01-2023	22-06-2023	Italy	44	Female	Smart TV	1 Month
11	Basic	10	16-05-2022	22-06-2023	United Sta	31	Female	Smartphone	1 Month
12	Premium	15	25-03-2023	28-06-2023	Canada	45	Male	Tablet	1 Month
13	Standard	12	30-11-2021	27-06-2023	United Kin	48	Female	Laptop	1 Month
14	Basic	10	01-08-2022	26-06-2023	Australia	27	Male	Smartphone	1 Month
15	Standard	12	09-05-2023	28-06-2023	Germany	38	Female	Smart TV	1 Month
16	Premium	15	07-04-2023	27-06-2023	France	30	Male	Tablet	1 Month
17	Basic	10	24-01-2022	25-06-2023	Brazil	30	Female	Laptop	1 Month
18	Standard	12	18-10-2021	24-06-2023	Mexico	63	Male	Smartphone	1 Month
19	Premium	15	15-02-2023	23-06-2023	Spain	42	Female	Smart TV	1 Month
20	Basic	10	27-05-2023	22-06-2023	Italy	41	Male	Tablet	1 Month
21	Premium	15	10-06-2023	22-06-2023	United Sta	26	Female	Laptop	1 Month
22	Basic	10	22-07-2022	28-06-2023	Canada	34	Male	Smartphone	1 Month
23	Standard	12	05-12-2021	27-06-2023	United Kin	49	Female	Smart TV	1 Month
24	Standard	12	03-04-2022	26-06-2023	Australia	31	Male	Tablet	1 Month
25	Basic	10	16-03-2023	28-06-2023	Germany	40	Female	Laptop	1 Month
26	Premium	15	12-01-2020	27-06-2023	France	29	Male	Smartphone	1 Month
27	Basic	10	29-08-2022	25-06-2023	Brazil	47	Male	Smart TV	1 Month
28	Standard	12	27-09-2021	24-06-2023	Mexico	33	Male	Tablet	1 Month

Netflix annual subscribers 2011

Date	Subscribers (mm)
2011	21.5
2012	25.71
2013	33.63
2014	47.98
2015	62.71
2016	79.9
2017	99.04
2018	124.35
2019	151.56
2020	182.86
2021	209
2022	220.6
2023	238.3

Netflix annual revenue 2011 to

Year	Revenue ($bn)
2011	3.1
2012	3.5
2013	4.3
2014	5.4
2015	6.7
2016	8.8
2017	11.6
2018	15.7
2019	20.1
2020	24.9
2021	29.6
2022	31.6

Purpose for choosing the dataset

In order to effectively analyse and get useful trends and patterns about How Netflix Movies & Tv Shows analysis Provided insightful information how audience votes and score impacts movies and tv shows ratings over the years, how movie duration time is changing over time. This research provides light on the various elements driving the popularity of content by carefully examining Number of votes received for each genre on IMDB scale and category preferences for Audience. This analysis makes producer's and movie production team to take more informed decisions about current audience preferences over Movies and Tv Shows.

Scope of work and BI questions

The questions this project seeks to answer are as follows:

- What content Netflix is adding over the years & which content holds majorly in Netflix -movies or Tv shows?
- Which countries produced most number of movies and TV shows in Netflix and percentage of movies and tv shows are produced for different age groups?
- Which month, Netflix adding most number of movies and tv shows?
- Movies run time preference by larger amount of audience and change of movie duration length over the years?
- Most popular movie and tv show genre over the years in terms of highest number of votes received and IMDB score?
- Most and least watched seasons for a tv shows by number of votes received?
- Monthly and yearly Revenue regenerated by subscription type of different countries?
- Netflix growth and downfall of subscribers over the years?

Data pre-processing and cleaning

Loading the data

The first step in this analysis is loading the dataset. This is done via get data option from home tab. This button gives a drop down presenting the different ways of data can be loaded into power bi.
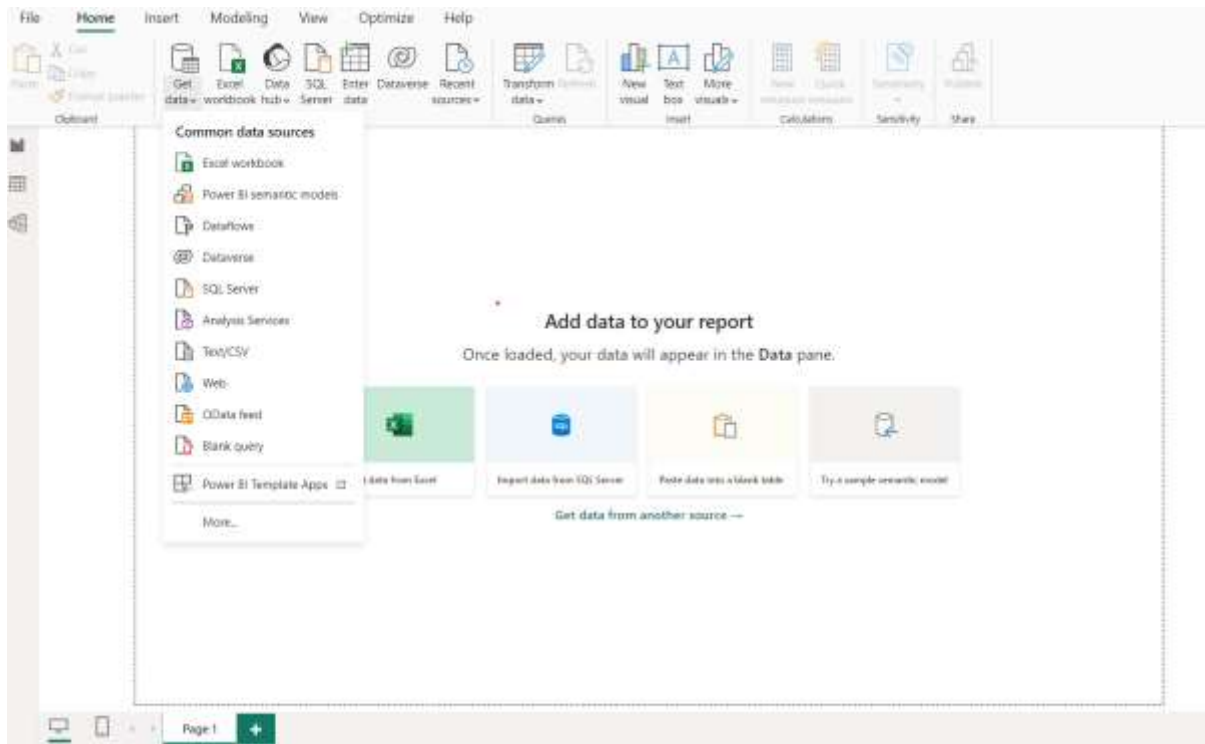
The data for this project is in csv format so the text /csv option was selected .

First the Netflix data set was chosen. This open the dialog box showing the top 200 rows of the data with an option to load , transform data or cancel .
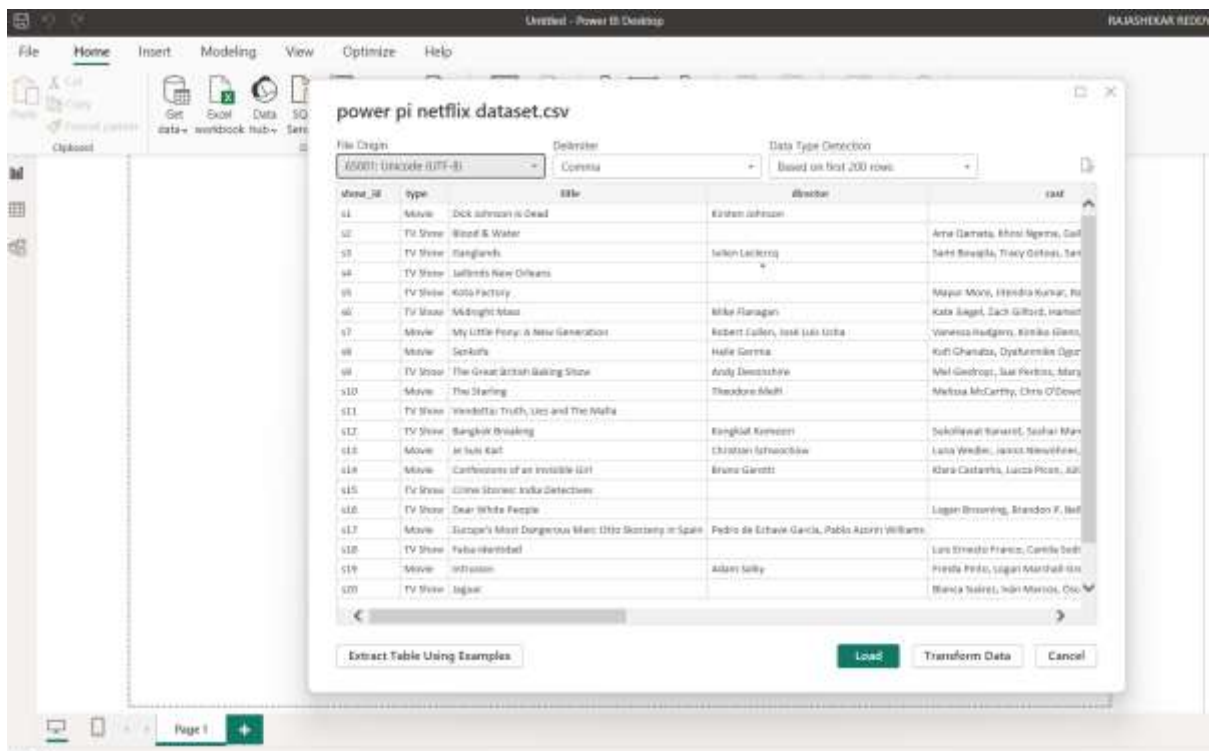
The same process was followed to load the other datasets into power bi.
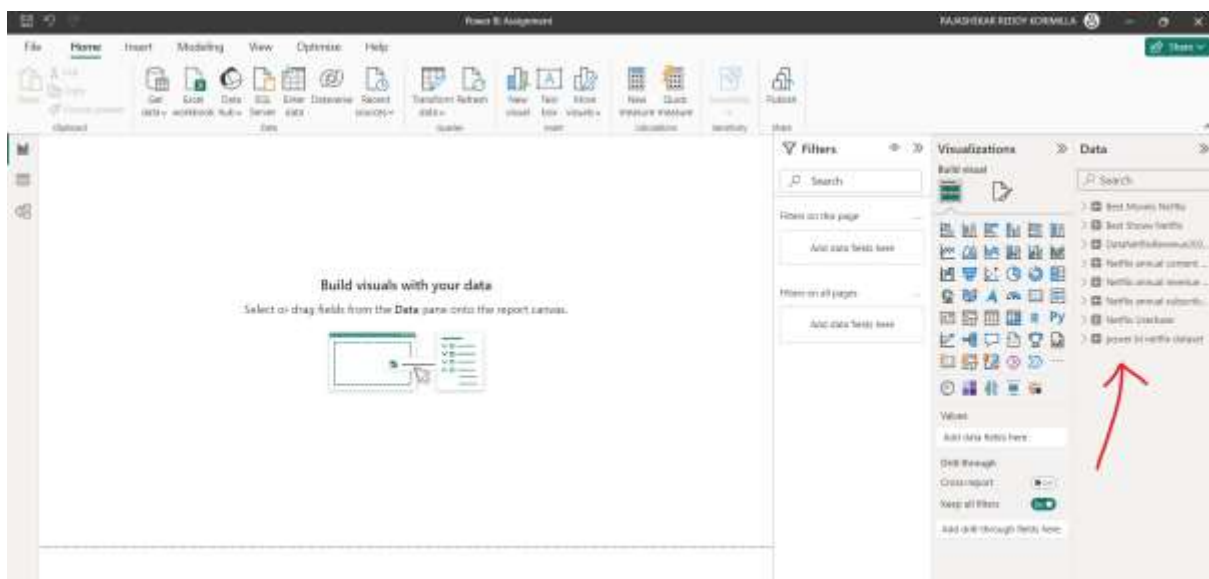
Data cleaning

The next step is the data cleaning and pre-processing which will be done in power query editor.

To perform this , use the option transform data , this will open power query editor

Removing unnecessary columns

Columns like show id , director, cast, description are not needed for analysis , it will be removed from the dataset.

Filtering Rows

The dataset contains unwanted rows in the rating column as shown in fig . so ww will filter out the unwanted rows from the rating column as shown in the fig.
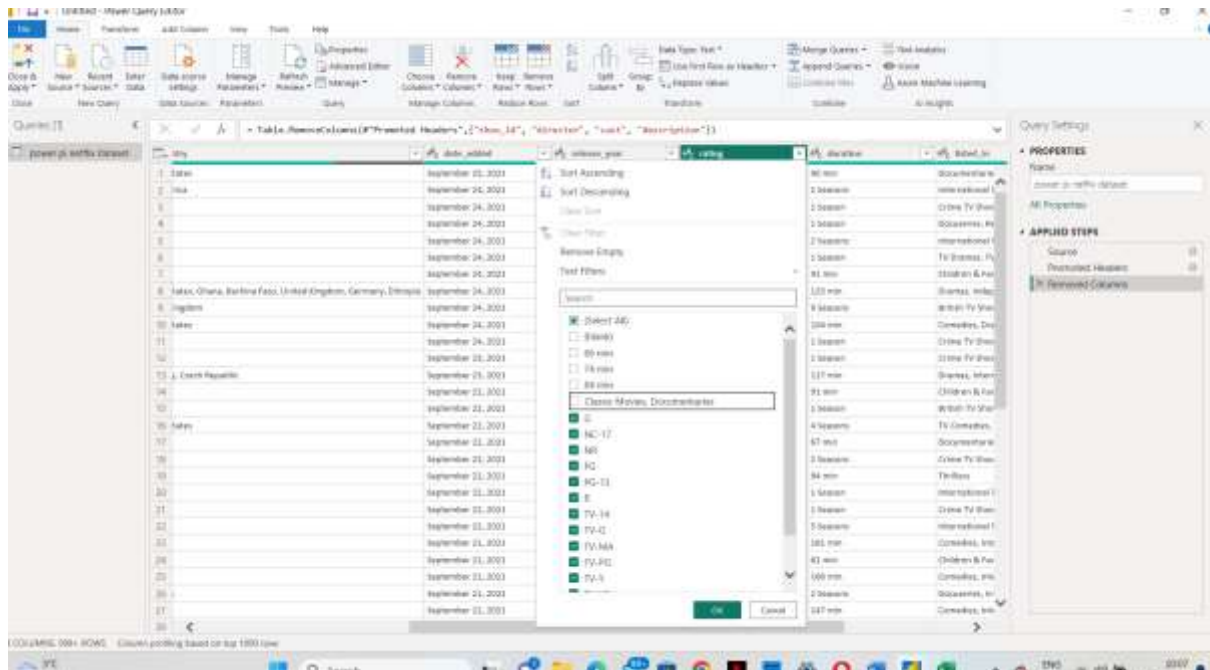
Removing Nulls /blanks

The country column contains a lot of blank rows and nulls rows which we have to deal with it.

`, each ([rating] <> "" and [rating] <> "66 min" and [rating] <> "74 min" and [rating] <> "84 min"`

	AᴮC country	AᴮC date_added	Aᴮ
	United States	September 25, 2021	2(
	South Africa	September 24, 2021	2(
		September 24, 2021	2(
		September 24, 2021	2(
	India	September 24, 2021	2(
		September 24, 2021	2(
		September 24, 2021	2(
	United States, Ghana, Burkina Faso, United Kingdom, Germany, Ethiopia	September 24, 2021	1!
	United Kingdom •	September 24, 2021	2(
	United States	September 24, 2021	2(
a		September 24, 2021	2(
		September 23, 2021	2(
	Germany, Czech Republic	September 23, 2021	2(
		September 22, 2021	2(
		September 22, 2021	2(
	United States	September 22, 2021	2(
o Skorzeny in Spain		September 22, 2021	2(
	Mexico	September 22, 2021	2(
		September 22, 2021	2(
		September 22, 2021	2(
ly Milligan		September 22, 2021	2(
	Turkey	September 22, 2021	2(
		September 21, 2021	1!
the Wheel		September 21, 2021	2(
	India	September 21, 2021	1!
	Australia	September 21, 2021	2(
		September 21, 2021	1!

we have many countries names with same names in country column, , so we need to remove the other rows and we will stick to only single country name.

Creating relationships

To be able to perform the analysis for this project , relationships need to be created between the tables in our model.

The first relationship is between Netflix dataset table and Best Movies Netflix best tv shows table based on the connecting column Content type and Movie Title this creates a relationship with many to one and a single cross-filter direction. This procedure is repeated for all other columns between these tables.

Next relationship is between Netflix annual revenue and Netflix revenue by area based on the connecting column revenue and year makes a one to many and a single cross-filter direction.

Next relationship is between Netflix annual revenue and Netflix revenue by area based on the connecting column revenue and year makes a one to many and a single cross-filter direction.

Next relationship is between Netflix annual subscribers and Netflix annual content spend by area based on the connecting column date and year makes a one-to-one relationship

BIG DATA AND BUSINESS INTELLIENCE MODULE

TOPIC - **NETFLIX MOVIES AND TV SHOWS ANALYSIS**

SECTION 2 : BUSINESS INTELLIGENCE SOLUTION

EXECUTIVE SUMMARY REPORT FOR STAKE HOLDERS

NAME: RAJASHEKAR REDDY KORIMILLA

Student ID-   W9633347

Submission Date- 10-05-2024

Executive summary

Introduction

Netflix is the world's largest online streaming platform with over 220 million subscribers in over 190 countries enjoying movies tv series, documentaries, feature films across a wide variety of genres and languages. around the world as of 2022. it important to see how effectively they merge the movies and tv shows on their platform in order to enhance the user experience, without losing subscribers churn.

This report seeks to answer the following questions

- What content Netflix is adding over the years & which content holds majorly in Netflix -movies or Tv shows?
- Which countries produced most number of movies and TV shows in Netflix and percentage of movies and tv shows are produced for different age groups?
- Which month, Netflix adding most number of movies and tv shows?
- Movies run time preference by larger amount of audience and change of movie duration length over the years?
- Most popular movie and tv show genre over the years in terms of highest number of votes received and IMDB score?
- Most and least watched seasons for a tv shows by number of votes received ?
- Monthly and yearly Revenue regenerated by subscription type of different countries?
- Netflix growth and downfall of subscribers over the years?

Data Set

This data set contains movies and tv shows from 1969-2022 including title release year, age , certification, runtime, genres, production countries, seasons information for tv shows, imdb score, number of votes, total revenue and subscribers.

Data model

The first relationship is between Netflix dataset table and Best Movies Netflix best tv shows table based on the connecting column Content type and Movie Title this creates a relationship with many to one and a single cross-filter direction. This procedure is repeated for all other columns between these tables.

Next relationship is between Netflix annual revenue and Netflix revenue by area based on the connecting column revenue and year makes a one to many and a single cross-filter direction.

Next relationship is between Netflix annual revenue and Netflix revenue by area based on the connecting column revenue and year makes a one to many and a single cross-filter direction.

Next relationship is between Netflix annual subscribers and Netflix annual content spend by area based on the connecting column date and year makes a one-to-one relationship

Findings Based on Analysis and Evaluation

Different visualizations were created to answer the data analysis questions and are detailed below

- **Question 1-** What content Netflix is adding over the years & which content holds majorly in Netflix -movies or Tv shows?
- Percentage of movies and tv shows are produced for different age groups?
- Which month, Netflix adding most number of movies and tv shows?
- Change of movie duration length over the years?

The below dashboard shows the visualization of charts used the answer this question
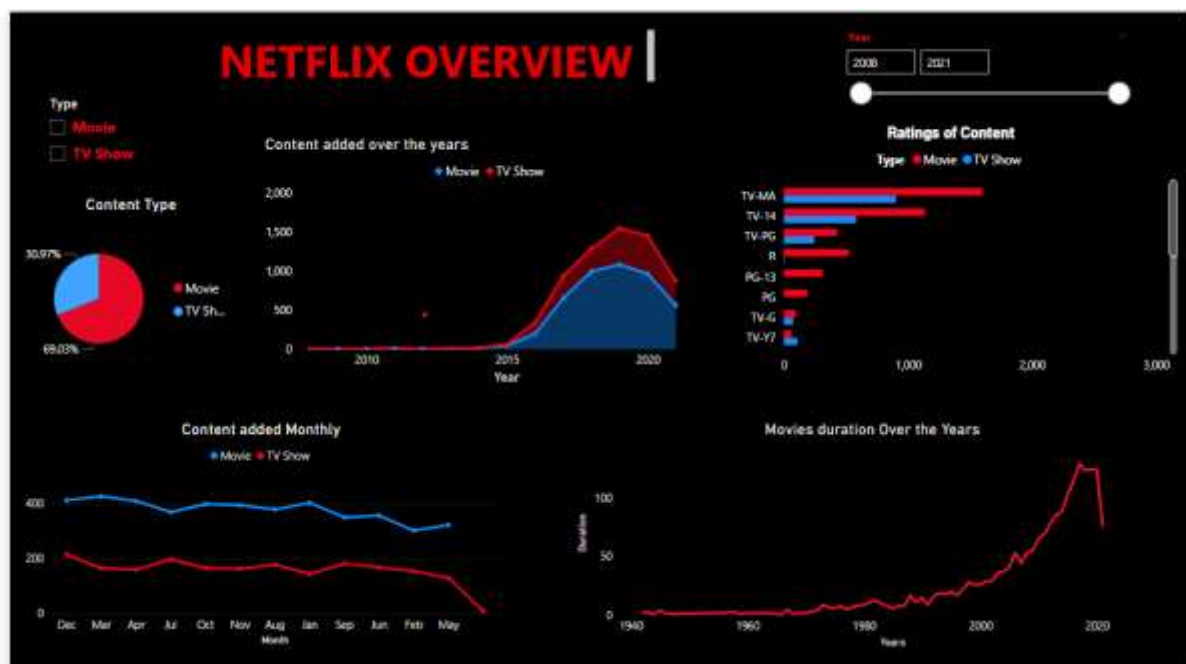
Two-line charts were created on this dashboard -one for monthly trends in total for movies and tv shows and other one shows the yearly trends of movie duration. Two slices were created -one for type of content to select and other one to set an interval for the years from 2008-2021. A play axis is added to see the change in trends across the years. A bar chart represents different Movie rating viewing count and area chart represents type of content added over the years.

Two measure was created using DAX formula for visualization on the dashboard

Month = FORMAT('power bi netflix dataset'[date_added],"MMM")

This measure is used to add extra month column from date column to show greater number of movies and tv shows added in which month

Day = FORMAT('power bi netflix dataset'[date_added],"DDD")

This measure is used to add extra month column from date column to show greater number of movies and tv shows added on the day of the week

Content By Type
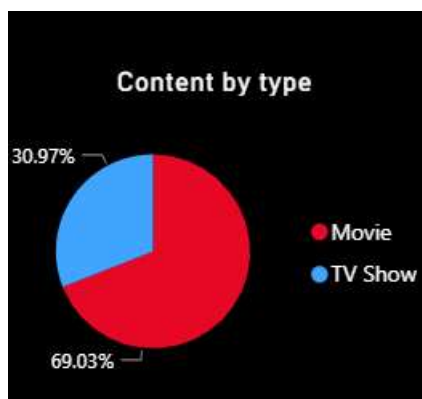
Pie chart represents the quantities of different constituent categories as a proportion of whole. The percentage values of each category add up to 100% and are represented by the angel of the wedge of pie.

This pie chart shows the total percentage of content between movies and tv shows over the years from 2008 to 2011.

2/3rd of the content on Netflix is movies and remaining 33% of them are tv shows. Two slices are used – one for type of content to show the interaction for charts and other allows to set an interval for the years

Content Added over the years
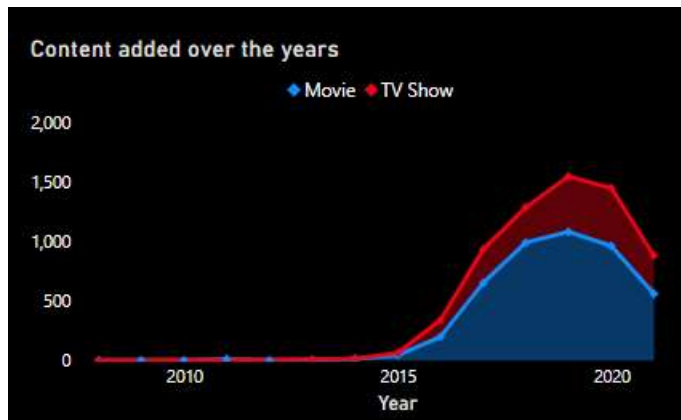
Area chart represents quantitates values for different categories over time to determine the trends and patterns to show how values change and fluctuate over time .

This area chart shows The growth in numbers of movies on Netflix is much higher then tv shows. About 2000 movies were added in both 2018 and 2019. The growth in content started from 2016 . There is decrease in the number of moves and tv shows added in the year 2020 which might be the reason of covid-19. And also there was a decrease in the number of movies added over the years , this shows that Netflix is focusing on adding more tv shows on its platform rather than movies
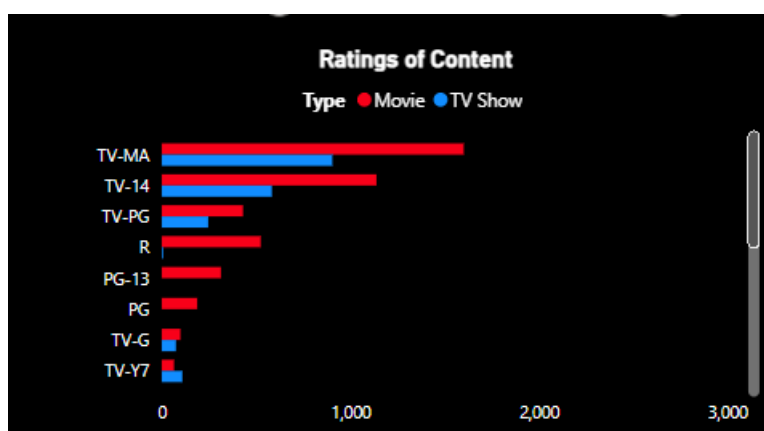
Bar chart represents values against the categories , bar length and height represents quantity , chosen a horizontal presentation for long categorical labels.

This bar chart shows, Around 50% of movies and tv shows are produced for adult audience i..e TV-MA, followed by young adults , older kids. Netflix has only few number of movies and tv shows for teenagers and other age groups.
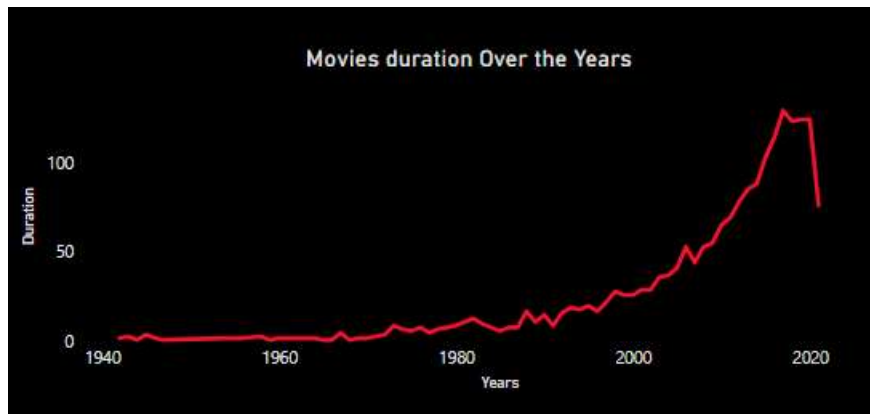
line represents quantitates values for different categories over time to determine the trends and patterns to show how values change and fluctuate over time.

This line chart shows the yearly trends of movie duration. From 1940 to till 2000, movie duration was very short which is below 50 Min but slowly it started increase from 2001 which goes up to 120 min. this shows how audience preferences over movie duration changing over the years
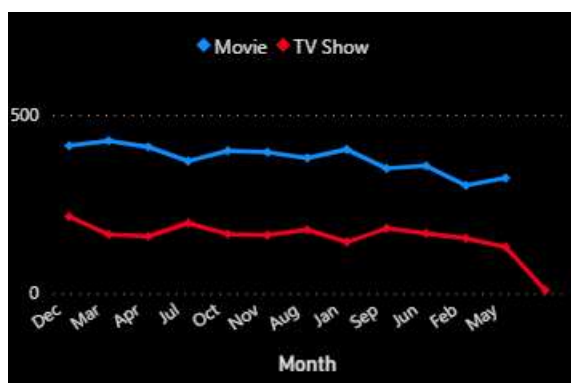
Line chart represents quantitates values for different categories over time to determine the trends and patterns to show how values change and fluctuate over time .

This graph illustrates Over the years from 2008 to 2022 , greater number of movies added in the month of mar December , April and jan while greater number of tv shows added in the month of December July September august.

- **Question 2**- Movies run time preference by larger amount of audience and Most popular movie genre over the years in terms of highest number of votes received and IMDB score ?

    The below dashboard shows the visualization of charts used the answer this question

This dashboard has clustered column chart to represents the values against major and minor categories, here more than 3 or 4 bars are cluttered together. There is Bar chart represents values against the categories, bar length and height represent quantity, chosen a horizontal presentation for long categorical labels. Line represents quantitates values for different categories over time to

determine the trends and patterns to show how values change and fluctuate over time. Two slices were created – one to see the duration of movie length and other one to set an interval for release years. A play axis is added to see the change in trends across the years. One KPI gauge used for total number of votes received.

One measure was created using DAX formula for visualization on the dashboard

Total Votes = sum('Best Movies Netflix'[NUMBER_OF_VOTES])

This measure is created to calculate total number of votes received for movies from 1969-2022 in Netflix
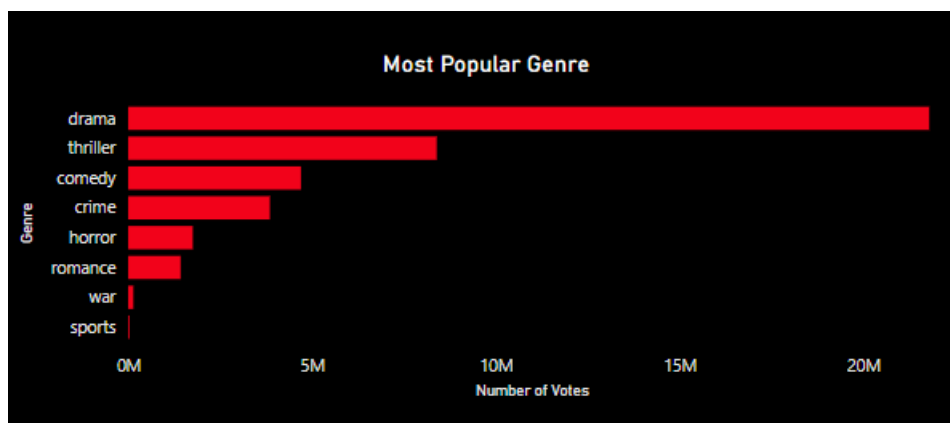
This graph illustrates, drama is the most popular genre in terms of votes received from audience followed by thriller and comedy genre. These three genres hold for about 50% of all the movies and tv shows. This value increase to 85% for top 7 genres.
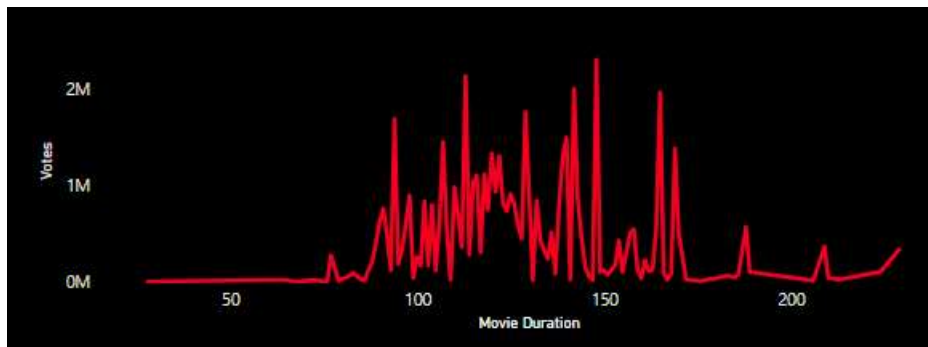
This graph illustrates, Highest number of votes received for movies from audience are among the duration of 90- 170 mins. Considering this fact that, larger amount of audience cannot watch movies which has more than 3 hours long in one sitting.
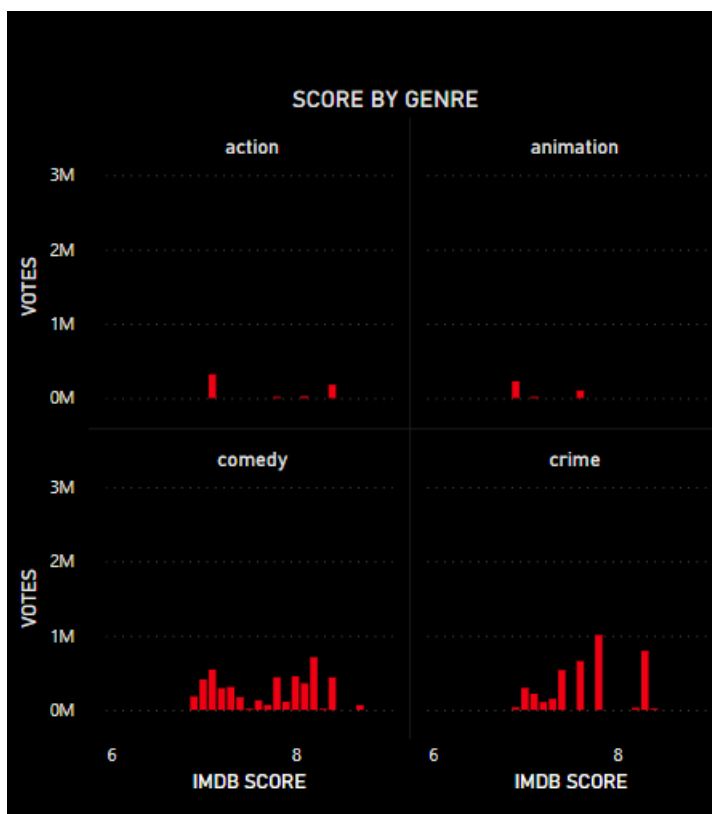
This graph illustrates, The drama, thriller are the most popular genre in terms of no of votes received from audience on IMDB scale followed by comedy and crime genre. These four genres hold for about 70% of all the movies.

Question 3 - Most popular tv show genre over the years in terms of highest number of votes received and IMDB score & Most and least watched seasons for a tv shows by number of votes received ?

- Which countries produced most number of Movies & TV shows in Netflix & TV Shows run time preference by larger amount of audience ?

The below dashboard shows the visualization of charts used the answer this question
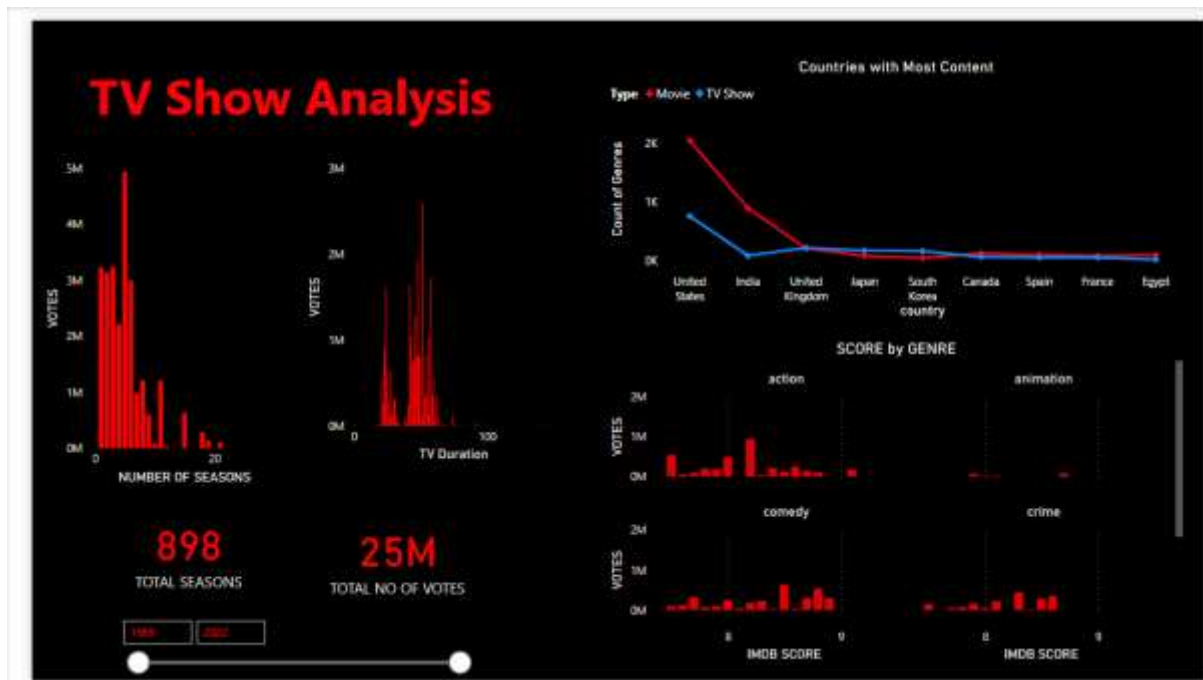
This dashboard has clustered column chart to represents the values against major and minor categories, here more than 3 or 4 bars are cluttered together. There are two Bar chart and one line represents values against the categories , bar length and height represents quantit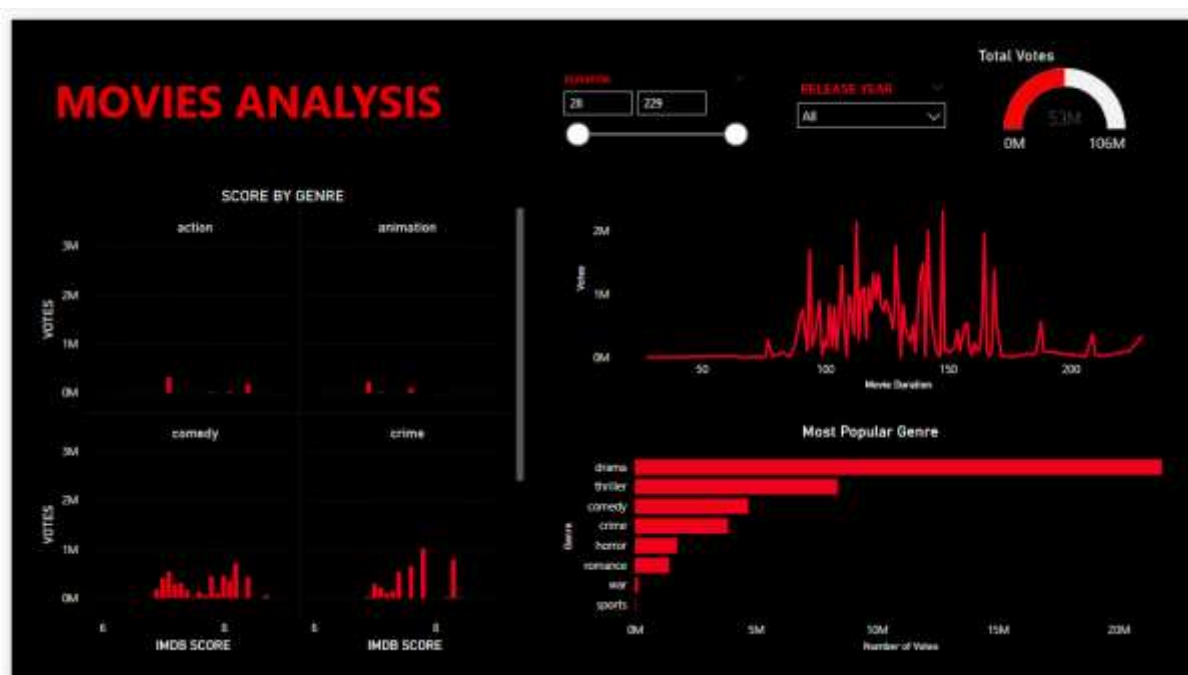y , chosen a horizontal presentation for long categorical labels. line represents quantitates values for different categories over time to determine the trends and patterns to show how values change and fluctuate over time. one slice were created to set an interval for release years. A play axis is added to see the change in trends across the years. Two cards visuals have been added to shows the total number of votes and total number of seasons for tv show.

Two measure was created using DAX formula for visualization on the dashboard

TOTAL NO OF VOTES = sum('Best Shows Netflix'[NUMBER_OF_VOTES])

This measure is used to calculate total number of votes received for all the shows in Netflix

TOTAL SEASONS = SUM('Best Shows Netflix'[NUMBER_OF_SEASONS])

This measure is used to calculate total number of seasons for all the shows in Netflix .
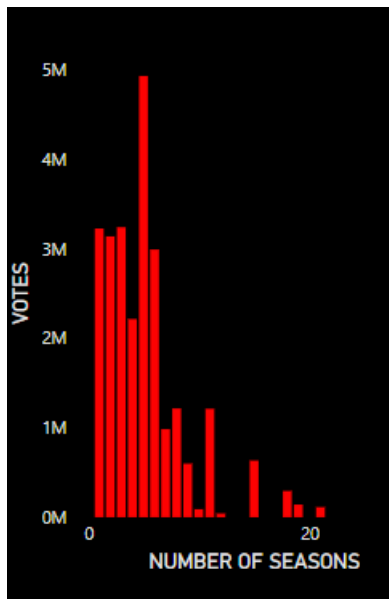
This graph illustrates Tv shows with five Seasons is most watched and has highest number of votes received from audience followed by season 1, 3, 2 and 6. after 6 seasons there been gradual decrease in watching the tv show, this shows larger amount of audience cannot watch tv shows which has more than 65-6 seasons.
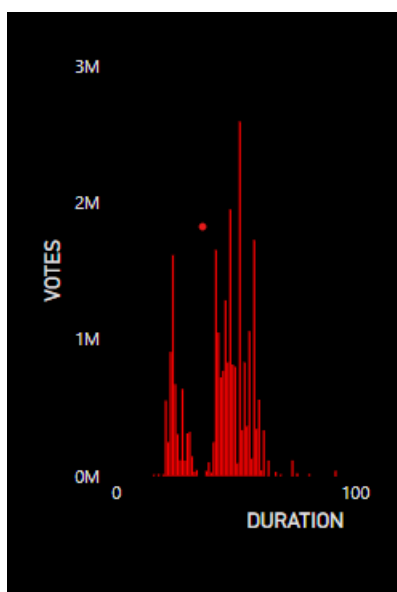
This graph illustrates ,Highest number of votes received for movies from audience are among the duration of 90- 170 mins. Considering this fact that , larger amount of audience cannot watch movies which has more than 3 hours long in one sitting.
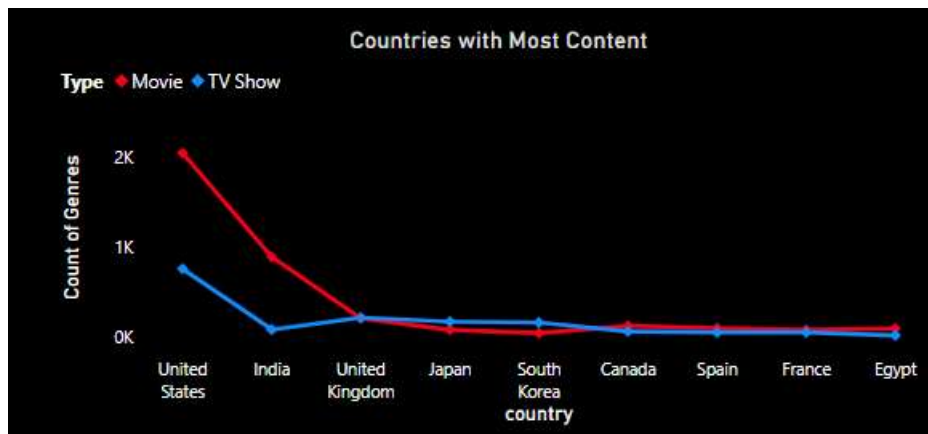

This graph illustrates, United sates produced highest number of tv shows & Movies Produced followed by India, United Kingdom and Japan.

This countries holds 80% of Movies and TV shows.
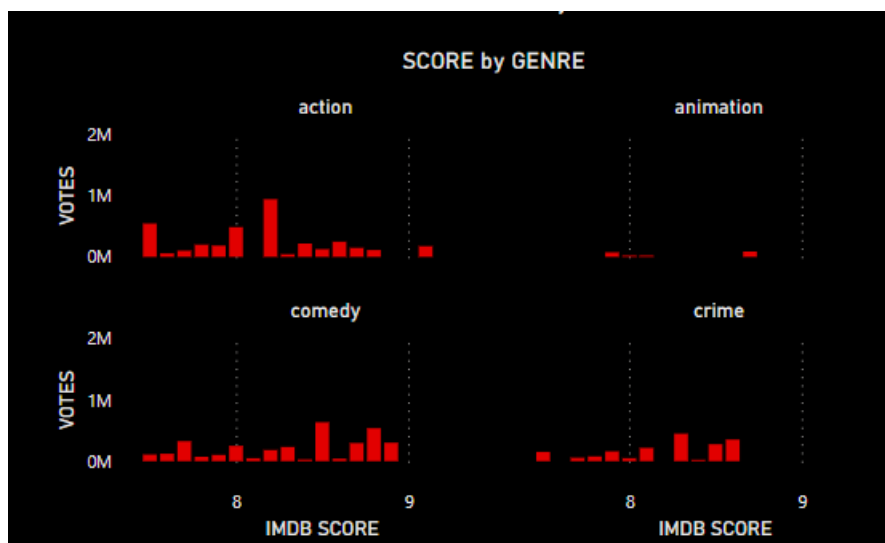
This graph illustrates, the drama, comedy, sci-fi are the most popular tv show genre in terms of no of votes received from audience on IMDB scale followed by thriller and action genre. These five genres hold for about 70% of all the movies.

Question 4- Monthly and yearly Revenue regenerated by subscription type of different countries & Netflix growth and downfall of subscribers over the years?

This dashboard has clustered column chart to represents the values against major and minor categories, here more than 3 or 4 bars are cluttered together. There are one Bar chart and one line and one area chart represents values against the categories, bar length and height represents quantity, chosen a horizontal presentation for long categorical labels. line represents quantitates values for different categories over time to determine the trends and patterns to show how values change and fluctuate over time. two slices were created to see the different subscription types, and other one to differentiate the gender male and female.

Two Measure was created with DAX Formulas

One measure is to show the total subscribers from 2011 to 2022 for Netflix

And other is to show total Revenue from 2011 to 2022

```
Total Subscribers = SUM('Netflix annual subscribers 2011'[Subscribers ])
```

```
Total Revenue = SUM('Netflix annual revenue 2011 to'[Revenue ($bn)])
```

Monthly Revenue by Country and Subscription Type
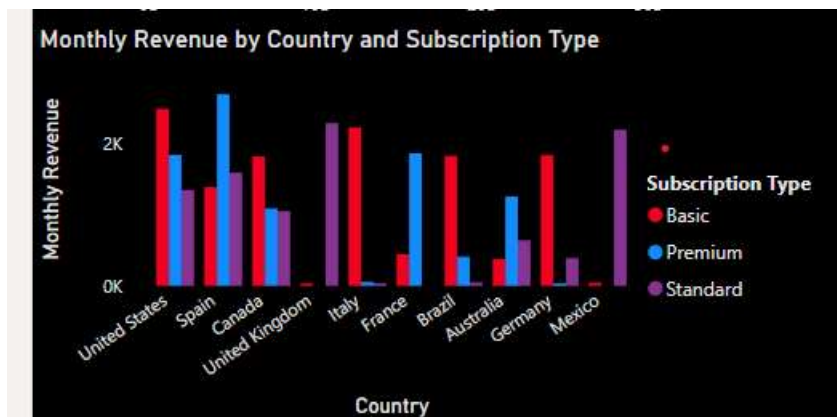
This graph illustrates United states and Italy has highest revenue generated by monthly basis for basic subscription type followed by Germany , Brazil , Canada ,while Spain has highest revenue generated in a month for premium subscription type, followed by France and united states,

And lastly united kingdom and Mexico has highest revenue generated in a month for standard subscription type followed by Spain and united states.

United states , united kingdom, Spain hold major revenue generated for Netflix online streaming service in a month for each subscription type.

This graph illustrates Over last 2.5 years, from 2018 Q1 to 2020 Q2 , united states and Canada , Europe and middle east Africa generated revenue over 1bn while Latin America and Asia pacific countries shown a stable revenue over the last 2 years .
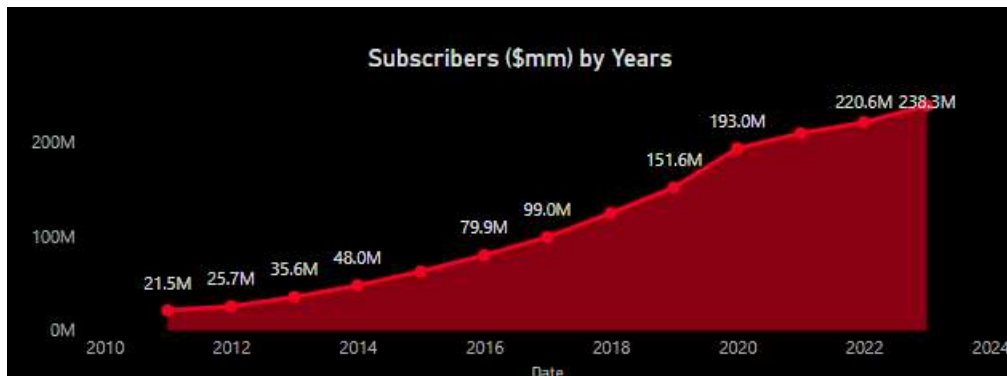
This graph illustrates Netflix growth of subscriber started steadily over the years till 2018, but from 2019-2022 , Netflix added 1bn new subscribers in total which is the tremendous growth for Netflix . covid-19 might be the reason to start adding new subscribers.

Key Findings

The following are the findings from the the business intelligence reports analysis of the above-mentioned business questions

- 2/3rd of content on Netflix is movies and remaining are tv shows
- Growth in adding number of movies is more on Netflix then Tv shows over the years
- United sates produced more number of movies and tv shows over the years than other countries
- Netflix produce large of number of movies and tv shows for adult audience
- December month is the most popular month for adding more number of movies and Tv shows
- 100-130 movie duration is most watched run time preference by larger amount of audience
- Drama is the most popular genre for movies and tv shows compared to other genres in terms of highest number of votes received on IMDB scale
- Tv show with Seasons 5 is most watched in terms of highest number of votes received from audience on IMDB scale
- United sates have highest revenue growth for Netflix over the years
- 2019-2022 years added up to 1 bn subscribers in total.

Recommendations

- Although Netflix hold highest % of movies but there has been gradually decreases in adding movies from 2020-2022 so Netflix should focus on adding more number of movies of different genres to avoid losing subscribers
- Only United states and India and United Kingdom added most of the movies and tv shows in Netflix, it should bring more movies and tv shows from different countries to make culture diversity and improve audience preference recommendation system.
- Movie and tv show runtime getting shorter over the years, Netflix should go with the larger amount of audience run time preference to have better viewing experience.
- Drama was the most popular genre with highest number of votes received in Netflix, with audience preferences changing over time, Netflix should bring other genres with good content to increase the userbase and not to stick with few genres.
- Only few countries having more accessible to Netflix streaming services like united sates, India and United Kingdom, to increase the more revenue and subscribers, Netflix should implement more accessible options of its streaming service to different parts of the world.

Conclusion

In conclusion, Netflix Movies & Tv Shows analysis Provided insightful information how audience votes and score impacts movies and tv shows ratings over the years, how movie duration time is changing over time. This research provides light on the various elements driving the popularity of content by carefully examining Number of votes received for each genre on IMDB scale and category preferences for Audience. The project emphasizes the value of data-driven decision-making in content strategy optimization by utilizing Microsoft Power BI for visualization. The results show that ongoing analysis is essential for maximizing viewer engagement and adjusting to changing viewer preferences. This initiative not only deepens our understanding about Netflix Movies and TV shows but also lays the groundwork for new discoveries, improved strategies, and well-informed choices for the producer's and movie production team to take

more informed decisions about current audience preferences over Movies and Tv Shows.