**Netflix IMDB Score Insights**

**Tools used** (Power BI, SQL, Python)

**Overview**:

The Netflix TV Shows and Movies dataset provides comprehensive information about various titles available on the popular streaming platform. The dataset includes details such as the title's name, its type (whether it is a TV show or a movie), a brief description of the content, the year it was released, age certification rating, runtime (for TV shows: length of episodes; for movies: duration), IMDb score, and IMDb votes.

**Ask**:

1. How is the IMDB score and votes are distributed to identify audience preferences?

**The problem(s) I am trying to solve:**

* Is there any relation between the insights so that producers and directors can identify the potential Movie and Show content to invest in.

**Business task**:

Analyze data to see how IMDB score relates with Age certification and help stakeholders to make informed decisions.

**Deliverable**:

To give recommendation(s) on which content to invest in for better engagement of audience.

### ****Prepare:****

**Data Location and Organization:**

I've downloaded the Netflix data set which contains content released from 1953 to 2023.

**Licensing** for use of data:

**I would like to thank and acknowledge for providing the dataset in CSV format.**

[Data\_source link](https://data.world/back2vizbasics%20%20)

**Answering my Question:**

I think the data can be aggregated and trends can be explored.

**Problems with the Data**:

The main problems with the data are numerous missing values in Age\_certification and for some shows the runtime is 0.

### ****Process****:

I chose Python to clean the data

**Loading data:**

**Python using Pandas Library**

**Cleaning Data**

# importing pandas for processing  
import pandas as pd  
  
#loading file to dataframe  
df = pd.read\_csv("D:/Data Analyst Projects/NetFlix\_Project/Netflix TV Shows and Movies.csv")  
  
#viewing records to have an idea  
print(df.head())  
  
print(df.describe())  
  
# check for null values in records  
print(df.isnull().sum())  
  
# fill the null values with ffill method...  
  
df["age\_certification"]=df["age\_certification"].fillna(method='ffill')  
df["imdb\_votes"]=df["imdb\_votes"].fillna(method='ffill')  
  
#deleting runtime=0  
df.drop(df.index[(df["runtime"]==0)],axis=0, inplace= True)  
  
print(df.isnull().sum())  
  
#saving cleaned df as csv  
df.to\_csv('D:/Data Analyst Projects/NetFlix\_Project/netflix\_imdbscores.csv')

* I have the loaded cleaned dataset to the Power BI.

**Transformation in PowerBI:**

* I have found out count of movies and shows.
* Then I have summed the rating, score columns.

### ****Analyze****:

**Trends**:

Due to the emergence of technologies, netflix gained popularity and the count of movies and shows released has grown enormously.

Due to Covid most people use Online streaming services for entertainment so we can see huge surge in no of movies and shows during 2019 and 2020

**Analysis Summary:**

Most IMDB Score is between 6 to 7 range.

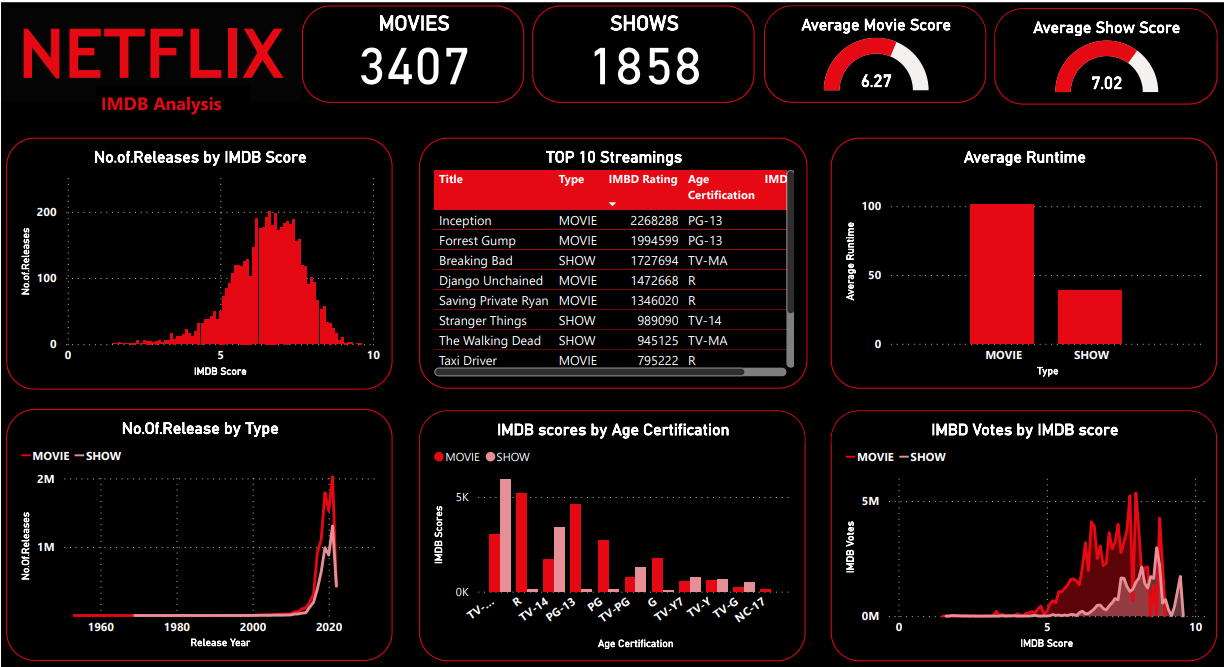
The highest show IMDB scores is 9.60 and lowest is 1.50.

The highest movie IMDB scores is 9 and lowest is 1.50.

Average movie score is 6.27 and for shows it is 7.02.

### ****Share****:

I have created the dashboard with Power BI, here is Snapshot of that.



### ****Act****:

**Conclusion:**

In my view Netflix can give recommendation for creators and producers to focus more on

* TV-14 and TV-MA for Shows since they have highest scores and ratings.
* R and PG-13 for Movies since they have highest scores and ratings.

Shows with 25 minutes to 50 minutes have highest IMDB ratings.

Shows are getting popularity in recent years so new producers can start with small shows and could scale up to movie production.