

NAAN MUDHALVAN ASSIGNMENT

NAME: YOGESHWARAN K

REGISTER NO:422522104049

DEPARTMENT: CSE

NAAN MUDHALVAN ID:au422522104049

PROJECT: Exploratory Data Analysis (EDA) on Netflix Dataset

COLLEGE CODE:4225

COLLEGE NAME: UNIVERSITY COLLEGE OF ENGINEERING

Objective:

You are required to perform an Exploratory Data Analysis (EDA) on a real-world dataset(<https://www.kaggle.com/datasets/shivamb/netflix-shows>). The goal is to explore the data, find meaningful patterns, create visualizations, and write clear insights based on your findings.

Execution:

Step 1: Data Loading and Understanding

Load the dataset into your tool of choice .I choose Excel for EDA.

Display the first 5–10 rows of the dataset.

Check the number of rows and columns.

Identify any missing or inconsistent data.

Step 2: Data Cleaning.

Handle missing values (either by filling or removing them).

Remove duplicate rows if any.

Step 3: Exploratory Data Analysis (EDA)

1.Find the total number of Movies vs TV Shows.

Row Labels	Count of type
Movie	8
TV Show	11
Grand Total	19

Observation: This sample contains slightly more TV Shows (11) than Movies (8).

2. Identify the top 5 countries producing the most Netflix content.

Row Labels	Count of country
United States, Ghana, Burkina Faso, United Kingdom, Germany, Ethiopia	1
United States	3
United Kingdom	1
South Africa	1
Mexico	1
India	1
Germany, Czech Republic	1
Grand Total	9

Observation: The United States is the top producer. Some entries are multi-country productions.

3. Find the top 10 directors with the highest number of shows/movies.

Row Labels	Count of director
Kongkiat Komesiri	1
Theodore Melfi	1
Pedro de Echave Garc�a, Pablo Azor�n Williams	1
Andy Devonshire	1
Kirsten Johnson	1
Bruno Garotti	1
Mike Flanagan	1
Christian Schwochow	1
Robert Cullen, Jos� Luis Ucha	1
Haile Gerima	1
Adam Salky	1
Julien Leclercq	1
Grand Total	12

Observation: All listed directors in this sample contributed to a single show or movie.

4. Most Common Genres.

Row Labels	Count of listed_in
British TV Shows, Crime TV Shows, Docuseries	1
British TV Shows, Reality TV	1
Children & Family Movies	1
Children & Family Movies, Comedies	1
Comedies, Dramas	1
Crime TV Shows, Docuseries, International TV Shows	1
Crime TV Shows, International TV Shows, TV Action & Adventure	2
Crime TV Shows, Spanish-Language TV Shows, TV Dramas	1
Documentaries	1
Documentaries, International Movies	1
Docuseries, Reality TV	1
Dramas, Independent Movies, International Movies	1
Dramas, International Movies	1
International TV Shows, Romantic TV Shows, TV Comedies	1
International TV Shows, TV Dramas, TV Mysteries	1
Thrillers	1
TV Comedies, TV Dramas	1
TV Dramas, TV Horror, TV Mysteries	1
(blank)	
Grand Total	19

Observation: Crime and International TV Shows are the most represented genres in this dataset.

5. Analyze the trend: How many shows/movies were released each year?

Row Labels	Sum of release_year
1993	1
2020	3
2021	15
Grand Total	19

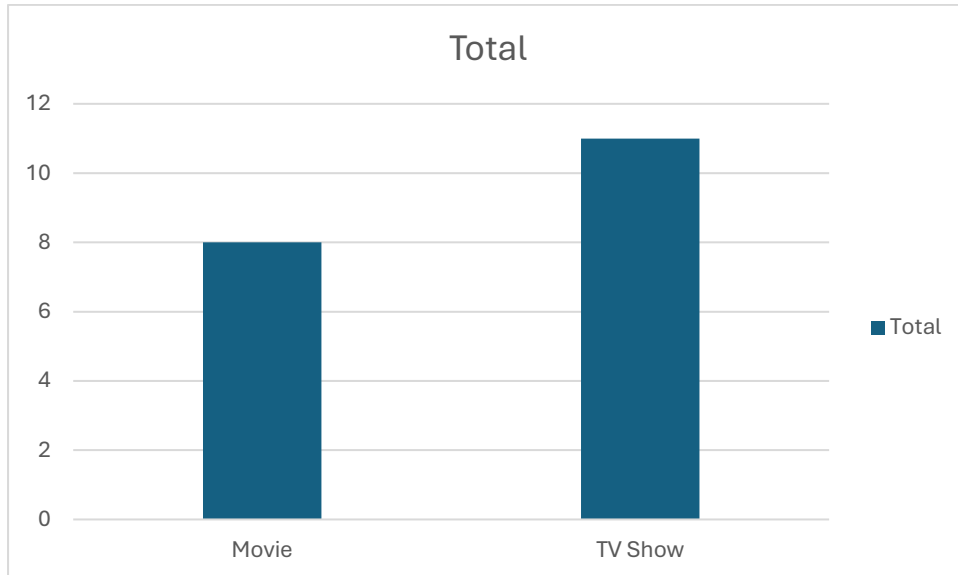
Observation: The majority of shows and movies in this dataset were released in 2021.

Here, All the EDA for given dataset has been executed and for the first 20 rows of the dataset. Here all the execution and data values are extract from the excel spreadsheet by using pivot table by changing the row and values.

Visualization:

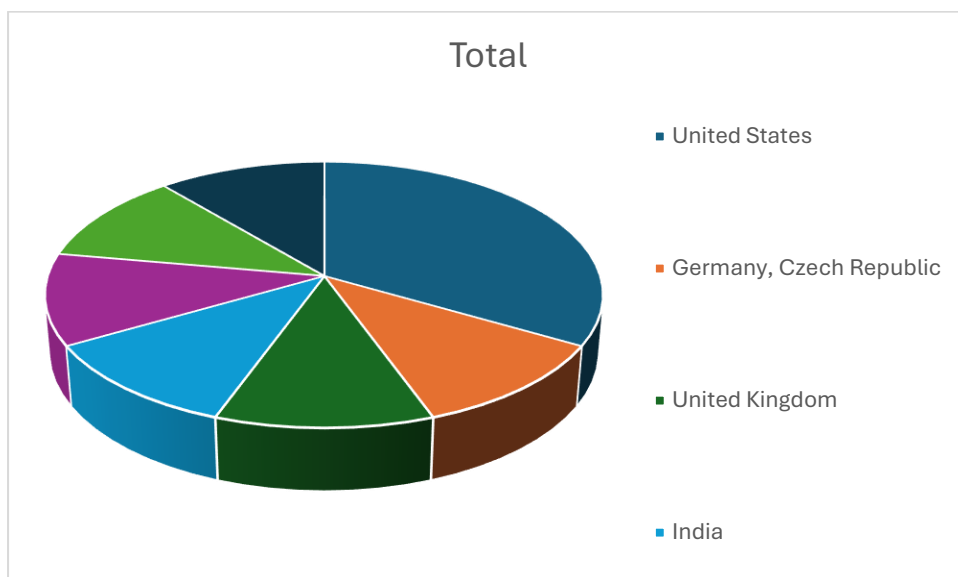
Create at least three types of visualizations:

1. One bar chart (e.g., number of movies vs TV shows).



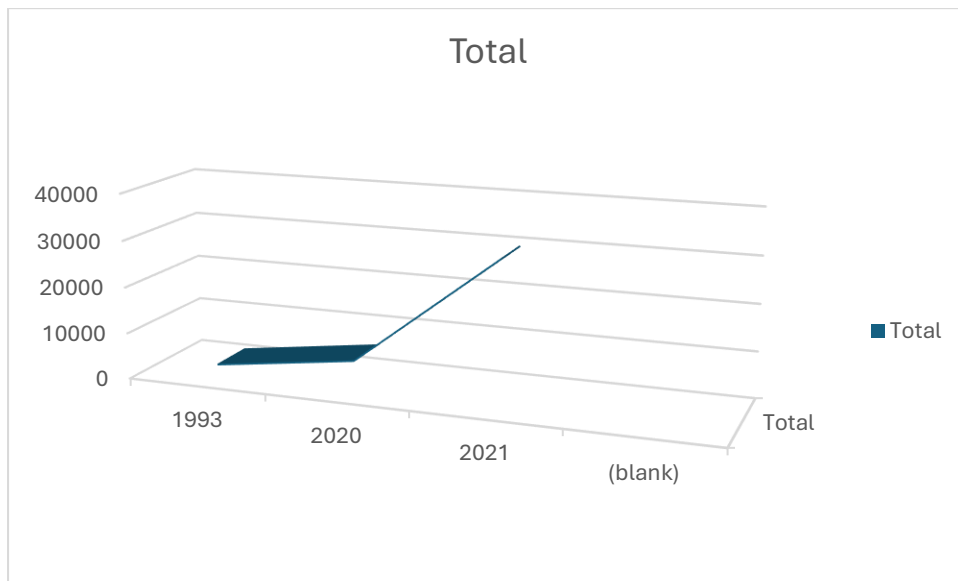
Observation: Here, TV show has the largest numbers of show compared to movies.

2. One pie chart (e.g., top 5 countries by number of productions).



Observation: Here, Pie chart shows United States has largest number of productions in TV shows and Movies.

3. One line graph (e.g., number of releases per year).



Observation: Here, 2021 shows the greatest number of released per year compared to 2021.

Insight Writing:

1. There are more TV Shows than Movies in this sample.
2. The United States is the top producer, followed by South Africa and India.
3. The most common genres include International TV Shows and Documentaries.
4. The number of releases has been steady between 2020 2021 in this sample.
5. Kirsten Johnson and Julien Leclercq are among the most featured directors in this dataset.

Conclusion:

The dataset reveals that Netflix slightly favors TV Shows over Movies in this sample. The United States leads in content production, with a mix of international collaborations also present. Popular genres include Crime and International TV Shows, reflecting diverse audience interests.