

# **RINEX MINI-PROJECT**

**COURSE: MACHINE LEARNING**

**TOPIC: EXPLORATORY DATA ANALYSIS**

NAME: ADHESH DHEERAJ

COLLEGE NAME: PES UNIVERSITY

BRANCH: CSE

SEMESTER - 3

EMAIL-ID: [adheshdheeraj21@gmail.com](mailto:adheshdheeraj21@gmail.com)

# Exploratory Data analysis:

## Inferences:

### Dataset – Netflix movies and TV shows

The screenshot shows a Jupyter Notebook interface with a code cell and its output. The code imports pandas as pd and reads a CSV file from the local path 'C:\Users\Dr.N.T.Murali Mohan\Desktop\netflixmovies\netflix\_titles.csv'. The output displays the first few rows of the dataset as a table.

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Gamata, Khosi Ngema, Gail Mabalane, Thabani...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...

#### 1) Total number of movies and TV shows:

The screenshot shows a Jupyter Notebook with a code cell that uses the value\_counts() method on the 'type' column of the DataFrame. The output shows there are 6131 movies and 2676 TV shows.

```
In [17]: df['type'].value_counts()
Out[17]: Movie      6131
         TV Show    2676
         Name: type, dtype: int64
```

#### 2) Number of movies and TV shows for each genre:

The screenshot shows a Jupyter Notebook with a code cell that uses the value\_counts() method on the 'listed\_in' column of the DataFrame. The output shows the count for various genres.

```
In [20]: df['listed_in'].value_counts()
Out[20]: Dramas, International Movies      362
         Documentaries                  359
         Stand-Up Comedy                 334
         Comedies, Dramas, International Movies  274
         Dramas, Independent Movies, International Movies  252
         ...
         Kids' TV, TV Action & Adventure, TV Dramas      1
         TV Comedies, TV Dramas, TV Horror              1
         Children & Family Movies, Comedies, LGBTQ Movies  1
         Kids' TV, Spanish-Language TV Shows, Teen TV Shows  1
         Cult Movies, Dramas, Thrillers                  1
         Name: listed_in, Length: 514, dtype: int64
```

3) Number of David Fincher movies

4) Number of R-rated movies

```
Name: listed_in, Length: 514, dtype: int64

In [50]: import numpy as np
ry=np.sum((df['director']=='David Fincher')&(df['type']=='Movie'))
ry

Out[50]: 6

In [43]: import numpy as np
r=np.sum((df['rating']=='R')&(df['type']=='Movie'))
r

Out[43]: 797
```

5) Number of Indian TV shows

```
In [47]: ind=np.sum((df['country']=='India')&(df['type']=='TV Show'))
ind

Out[47]: 79
```