EDA ON NETFLIX MOVIES AND TV SHOWS DATASET

Presented By : Harsh Mishra
Guide By : Naina Devi
College : Techno India University

AGENDA:

- To understand the composition and trends of Netflix’s content library.
- To analyze movies and TV shows based on country, genre, release year, and ratings.
- To visualize data insights through graphs and charts.
- To identify patterns in Netflix’s content strategy

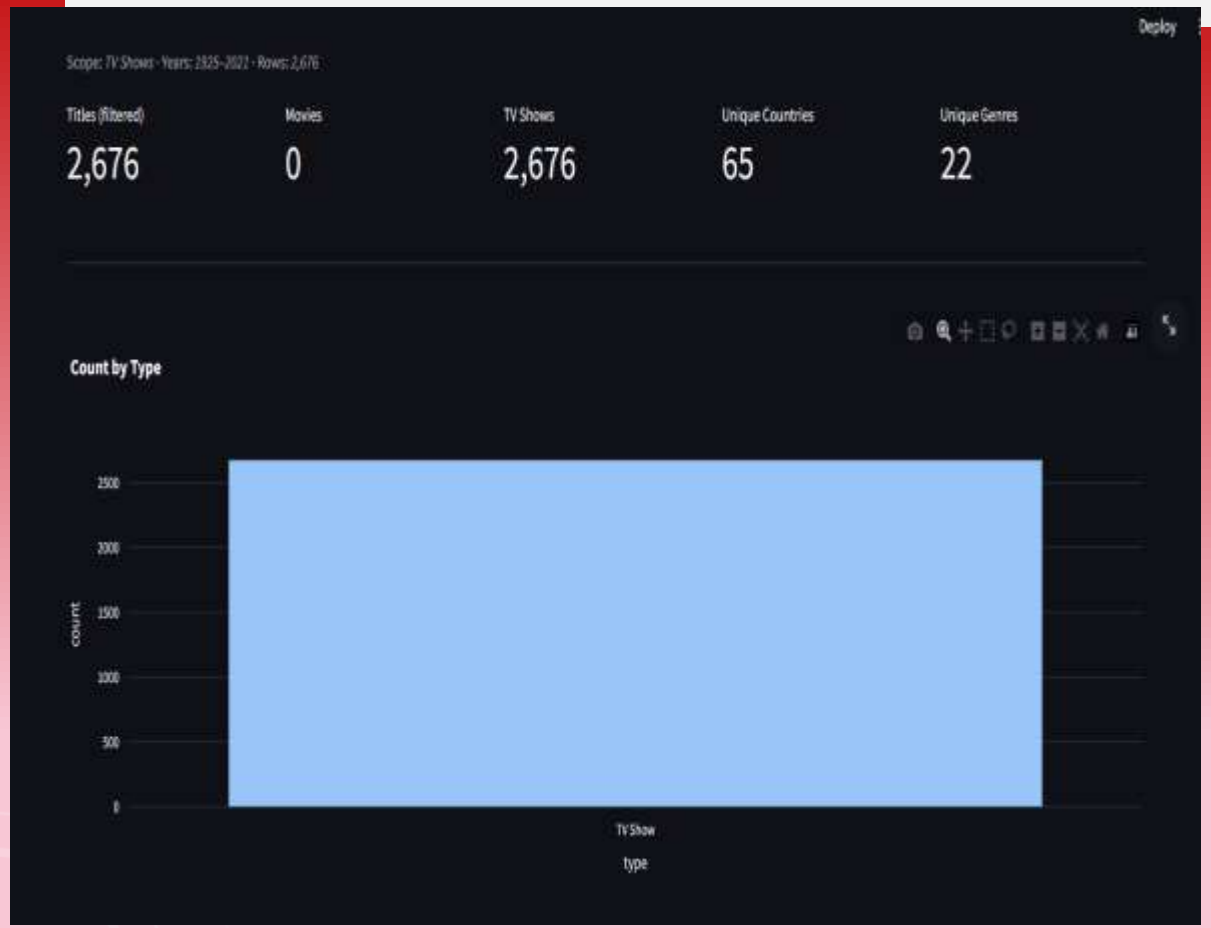
INTRODUCTION :

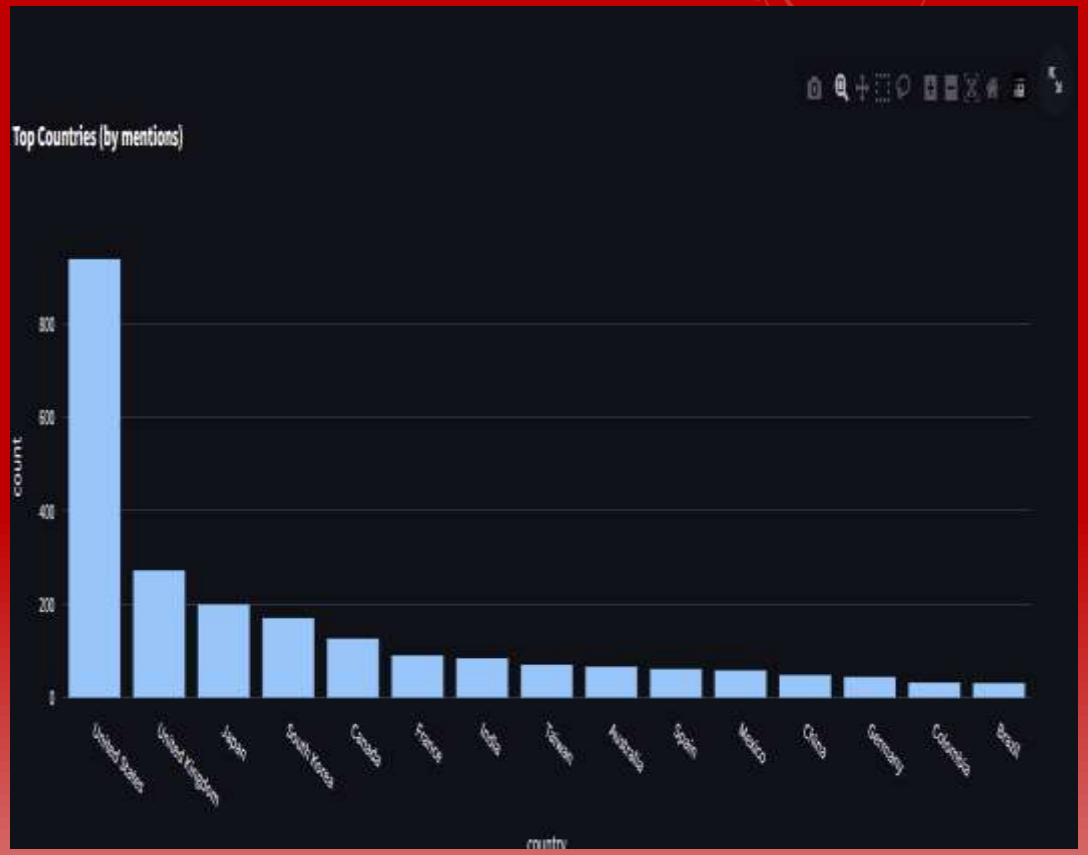
This project performs **Exploratory Data Analysis (EDA)** on Netflix's Movies and TV Shows dataset. The dataset includes information like title, director, cast, country, release year, rating, duration, and type of content. The goal is to gain meaningful insights into Netflix's content distribution and trends. Through data cleaning, visualization, and analysis, the project helps understand what kind of content Netflix produces most, how it has evolved over time, and which countries contribute the most to its library.

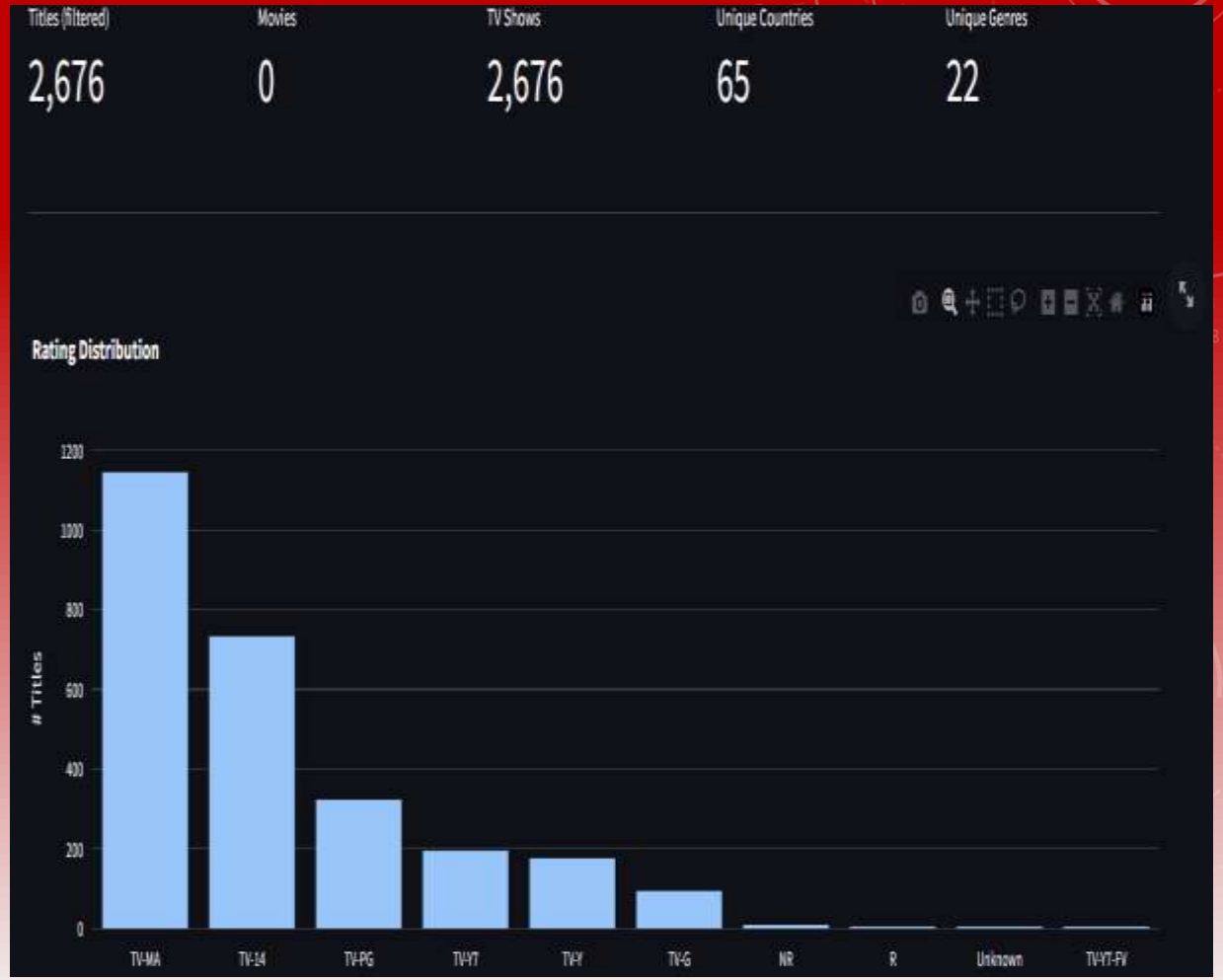
LIBRARIES AND TECHNOLOGIES USED

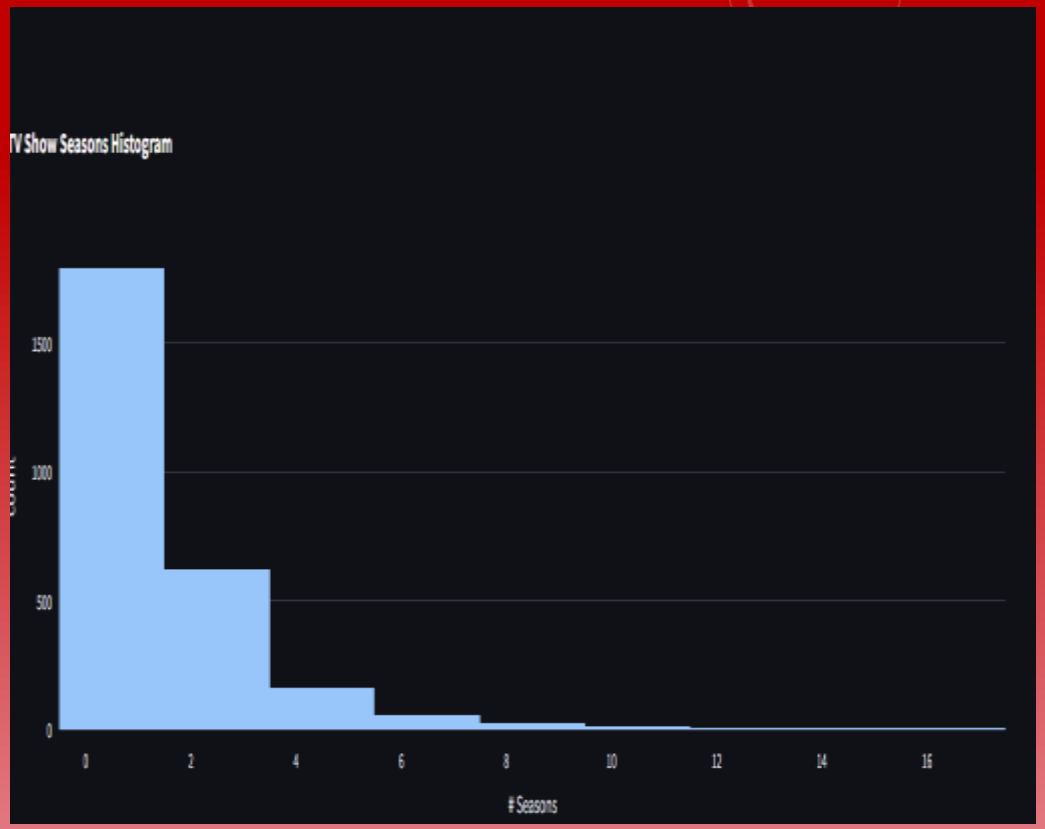
- **Programming Language:** Python
- **Libraries Used:**
  - **Numpy** for data manipulation
  - **Matplotlib**
  - **Sckit** learn for predictive modeling
  - **Joblib** for model serialization
- **Deployment and Interface:**
  - **Streamlit** for rapid development
  - **Render** for cloud deployment
- **Dataset Source:**
  - Kaggle
  - Mymoviedb.csv provides rich features for Netflix shows EDA
  - Development in **Jupyter Notebook** and version control with **GitHub**

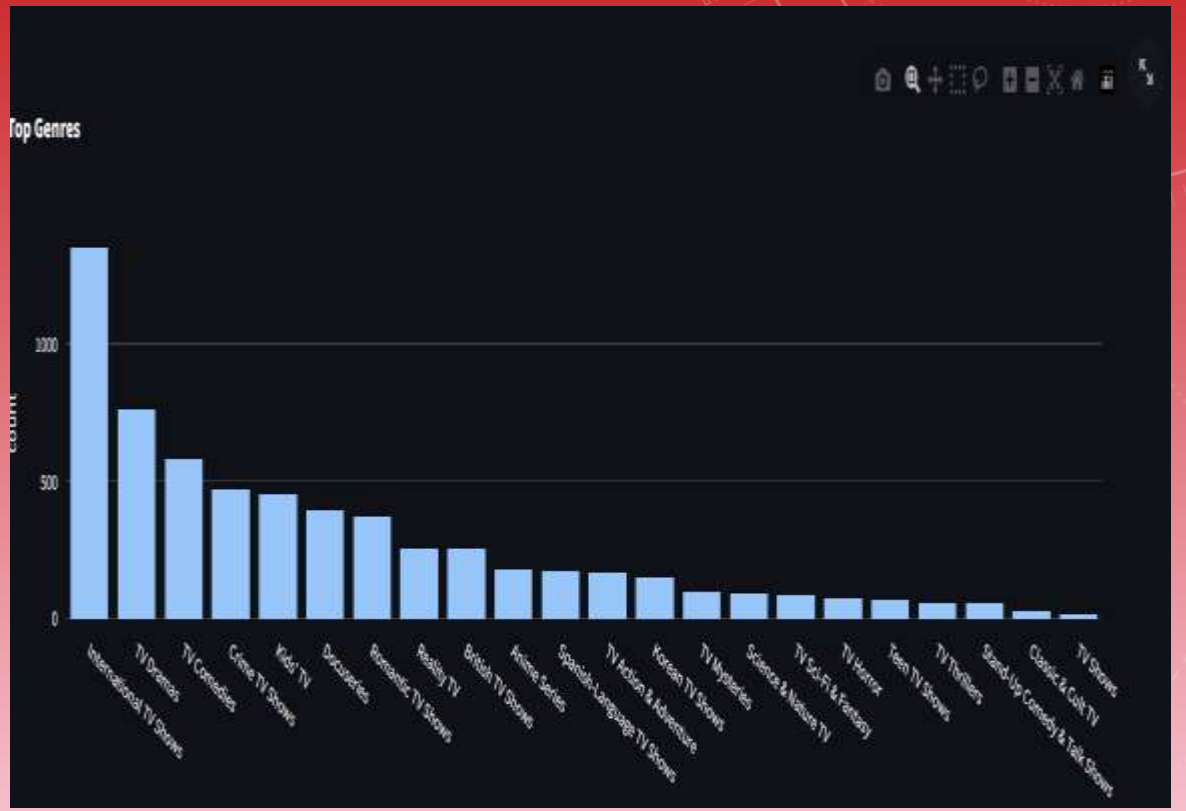
PROJECT WORKING:

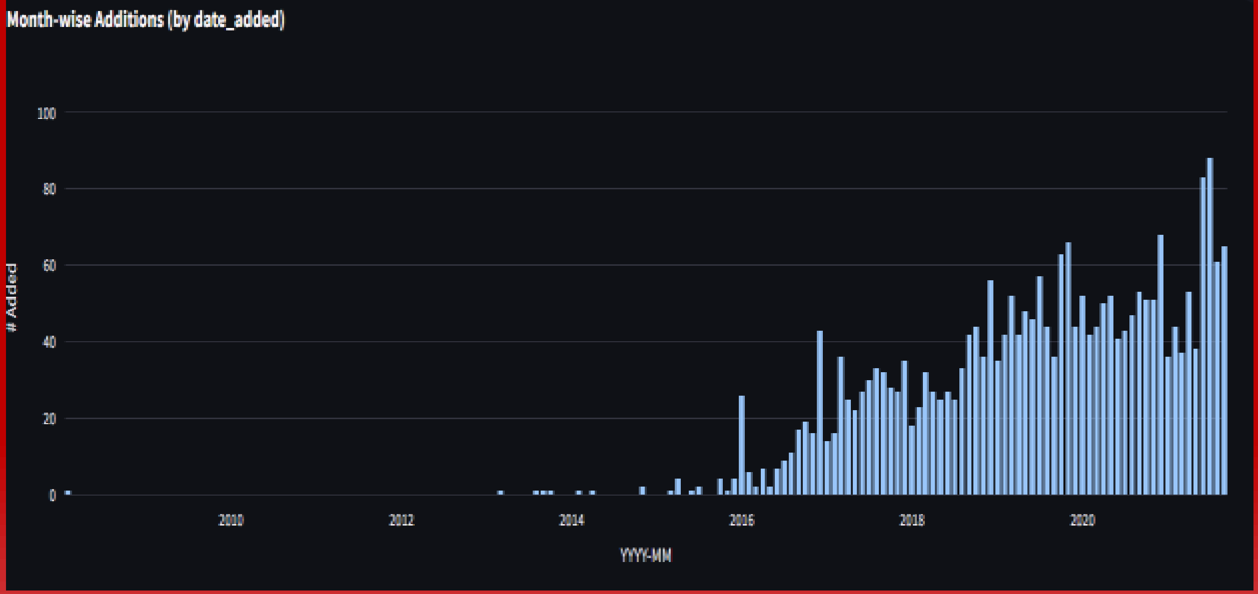



Top Countries (by mentions)
count
800
600
400
200
0
United States
United Kingdom
Japan
South Korea
Canada
France
India
Taiwan
Australia
Spain
Mexico
China
Germany
Colombia
Brazil
country

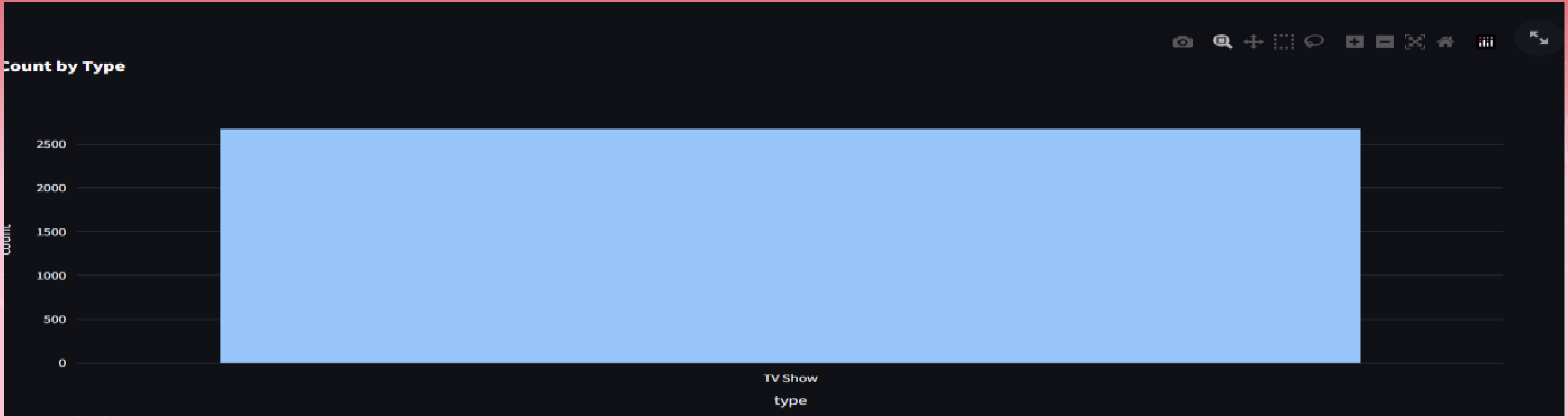

Titles (filtered)
2,676
Movies
0
TV Shows
2,676
Unique Countries
65
Unique Genres
22
Rating Distribution
Titles
1200
1000
800
600
400
200
0
TV-MA
TV-14
TV-PG
TV-Y7
TV-Y
TV-G
NR
R
Unknown
TV-Y7-FV


TV Show Seasons Histogram
1500
1000
500
0
0
2
4
6
8
10
12
14
16
Seasons


Top Genres
1000
500
0
International TV Shows
TV Dramas
TV Comedies
Crime TV Shows
Kids' TV
Documentaries
Romantic TV Shows
Reality TV
British TV Shows
Anime Series
Spanish-Language TV Shows
TV Action & Adventure
Korean TV Shows
TV Mysteries
Science & Nature TV
TV Sci-Fi & Fantasy
TV Horror
Teen TV Shows
TV Thrillers
Stand-Up Comedy & Talk Shows
Classic & Cult TV
TV Shows


Month-wise Additions (by date_added)
100
80
60
40
20
0
Added
2010
2012
2014
2016
2018
2020
YYYY-MM


Love
World
Life
Girl
La
Power
Rangers
Adventure
Secret
Time
Serie
Show
Crime
City
Black
Tale
Man
Day
Love
House
Stories
War
Killer
First
Night
Master
One
High
Inside
Murder
Kingdom
Friend
Monster
Dragon
King
Queen
Story
Good
Legend
Game
Last
Little
Blood
Happy
Home
Boy
Force
Planet
Family
Super
Transformers
Space
Marvel
Heart
Club
School
Lost
Magic
New
Rescue
Wild
American
Detective
LEGO
Kid
Dark
Made
Earth
Live
Call
Mind
Go
File
Bad
Great
Money
Age
Back
Big
Universe
Ghost
True
Baby
Song
Comedy
Nature
Next
Hotel
History
Fate
Squad
Word
Thing
Bangkok
Academy


Count by Type
2500
2000
1500
1000
500
0
count
TV Show
type

APPLICATIONS:

•Helps Netflix and other streaming services understand global content trends.

•Useful for data analysts and entertainment researchers.

n

•Supports decision-making in content acquisition and production.

•Educational use for learning data visualization and analytics.

PROBLEM FACED AND ITS SOLUTIONS:

CHALLENGES

Missing Values:
Many entries in *director*, *cast*, and *country* columns were missing.

Duplicate Records:
Some titles appeared more than once in the dataset.

Unformatted Dates and Durations:
The *date_added* and *duration* columns had inconsistent formats.

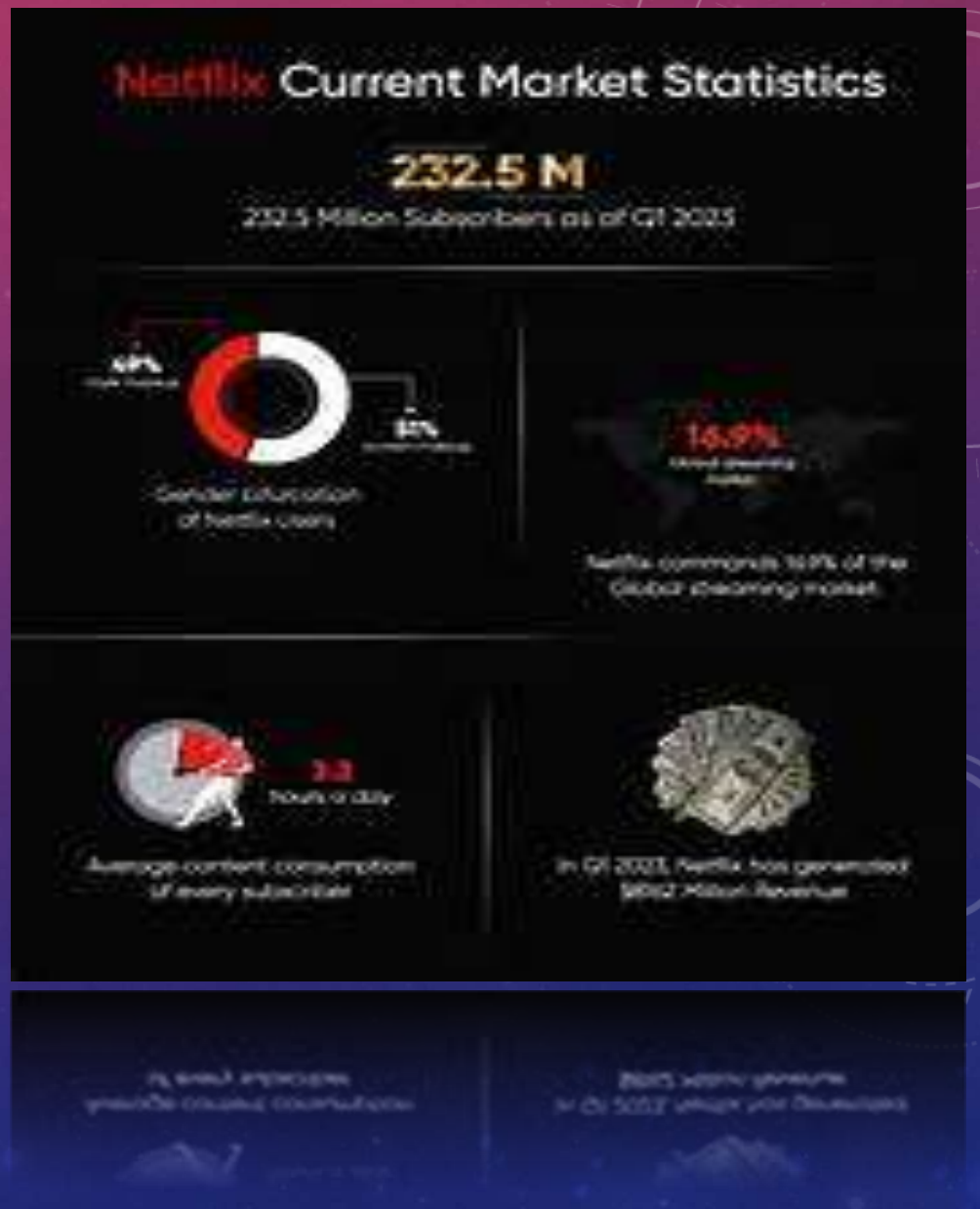
Cluttered Visualizations:
Graphs were overcrowded with too many categories.

SOLUTIONS

- Filled missing values with “Unknown” or removed incomplete rows using Pandas.
- Used drop_duplicates() to clean the data and ensure accuracy.
- Converted them to proper formats using pd.to_datetime() and string operations.
- Displayed only top 10 values for better clarity and presentation

FUTURE SCOPE:

- Extend analysis using machine learning to predict content popularity.
- Include IMDb ratings and viewer engagement data for deeper insights.
- Create an interactive dashboard using Power BI or Tableau.
- Automate updates with Netflix’s latest data through APIs.



PROJECT LINK :

- https://github.com/MishraHarsh25/EDA-on-Netflix

- https://eda-on-netflix-sgse.onrender.com/

THANK YOU