EDA ON NETFLIX MIOIES ON MOVIES AND TV SHOWS DATASET

- Presented By: **Shreya Singh**
- Guided By: Naina Devi
- College: Techno India University

AGENDA:

- Understand and analyze the Netflix movies and TV shows dataset.
- Perform data cleaning and preprocessing.
- Visualize important trends and insights.
- Solve encountered problems and discuss solutions.
- Explore potential applications and future development.

INTRODUCTION

This project explores the Netflix Movies and TV Shows dataset using Exploratory Data Analysis (EDA). The dataset contains details about various titles, including their type, cast, country, release date, and genre. By analyzing this data, we can uncover interesting trends and patterns in what's available on Netflix and better understand its content.

OBJECTIVE:

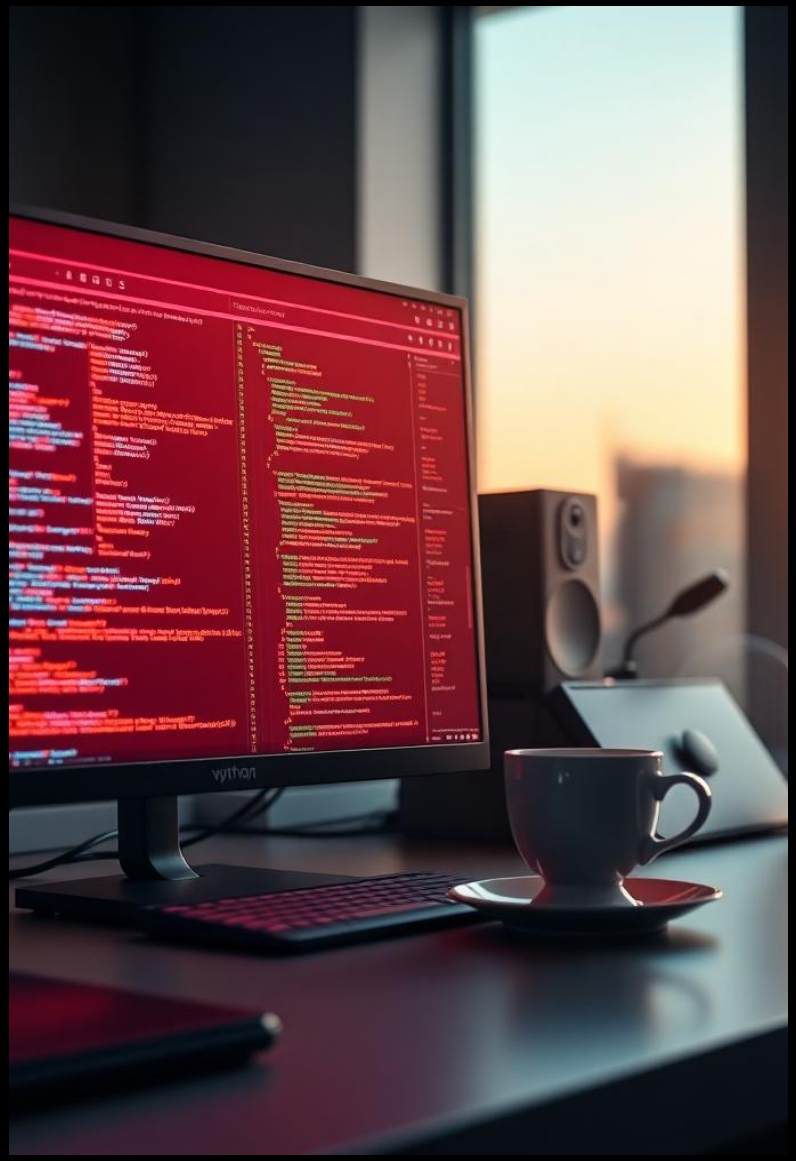
- The objective of this EDA is to examine the Netflix Movies and TV Shows dataset to find key patterns, trends, and insights about the content available on the platform
- To identify content trends and viewer preferences on Netflix to help inform decisions about what types of movies and shows to recommend or produce.

LIBRARIES AND TECHNOLOGIES USED:

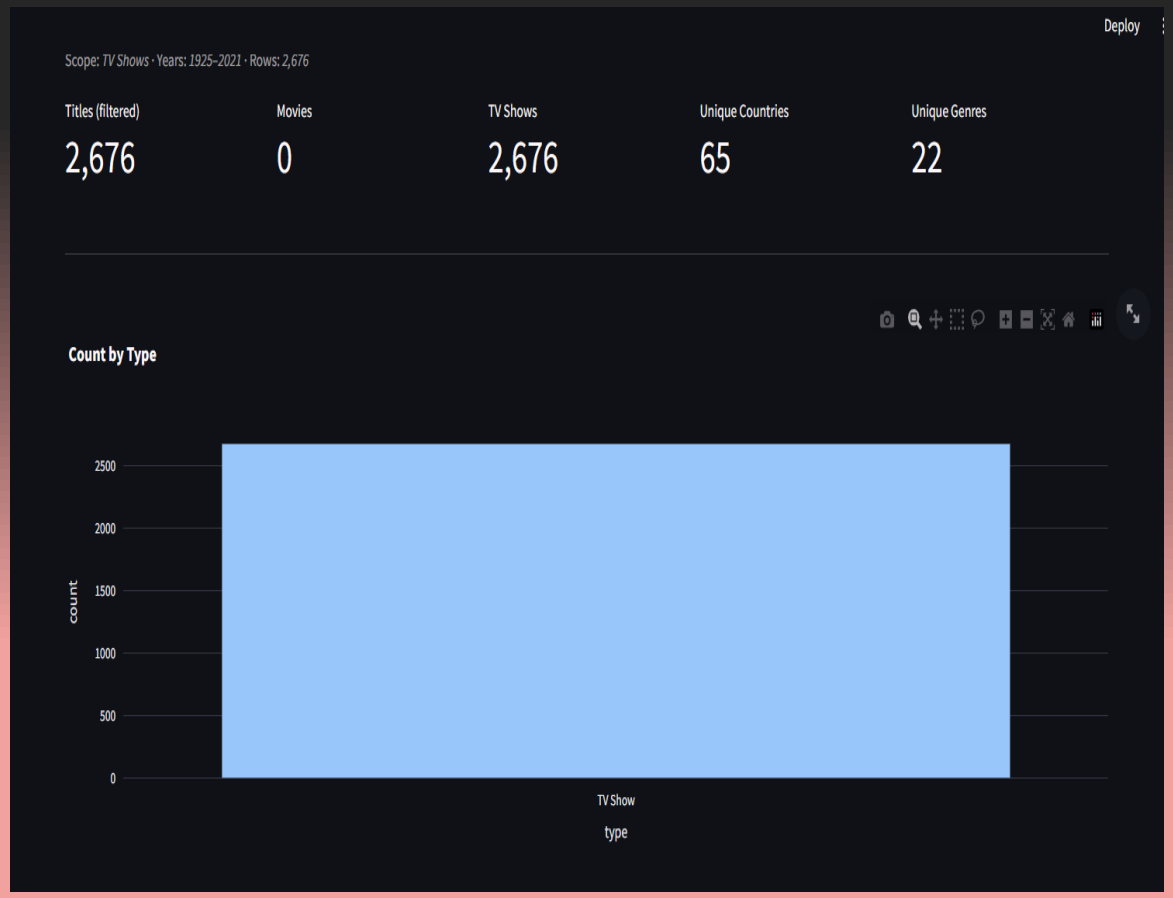
Programming Language: Python

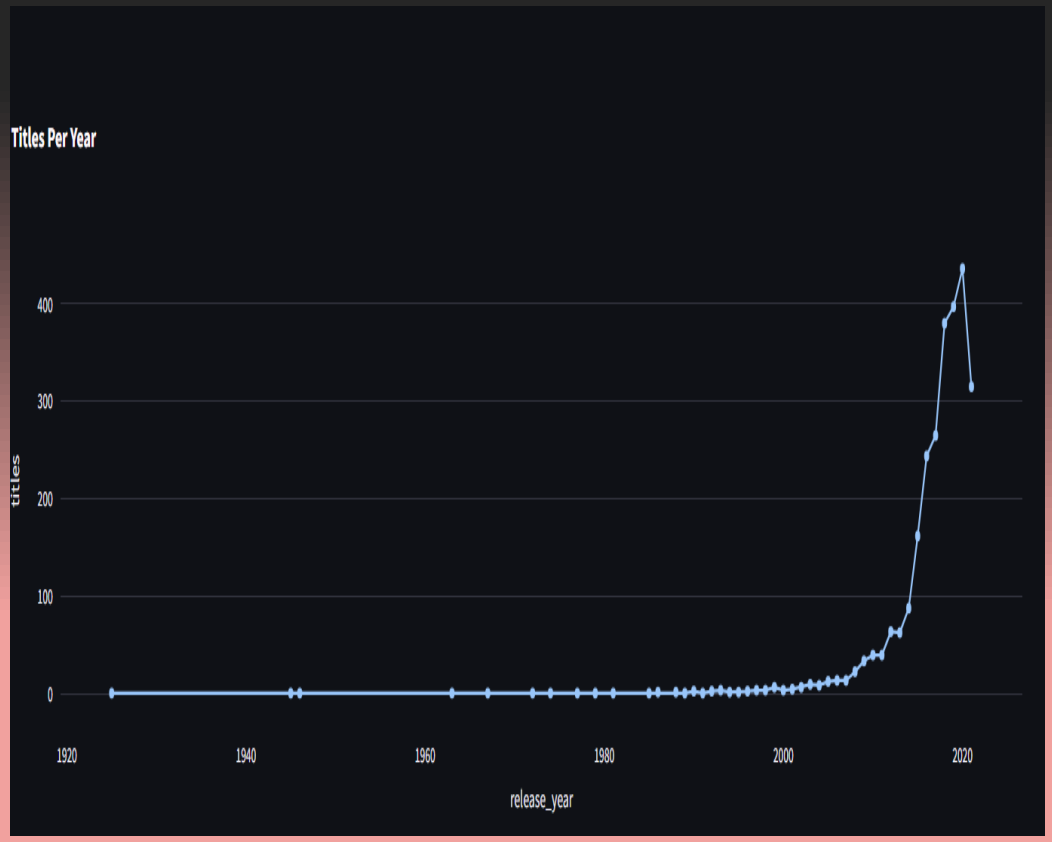
Libraries Used:

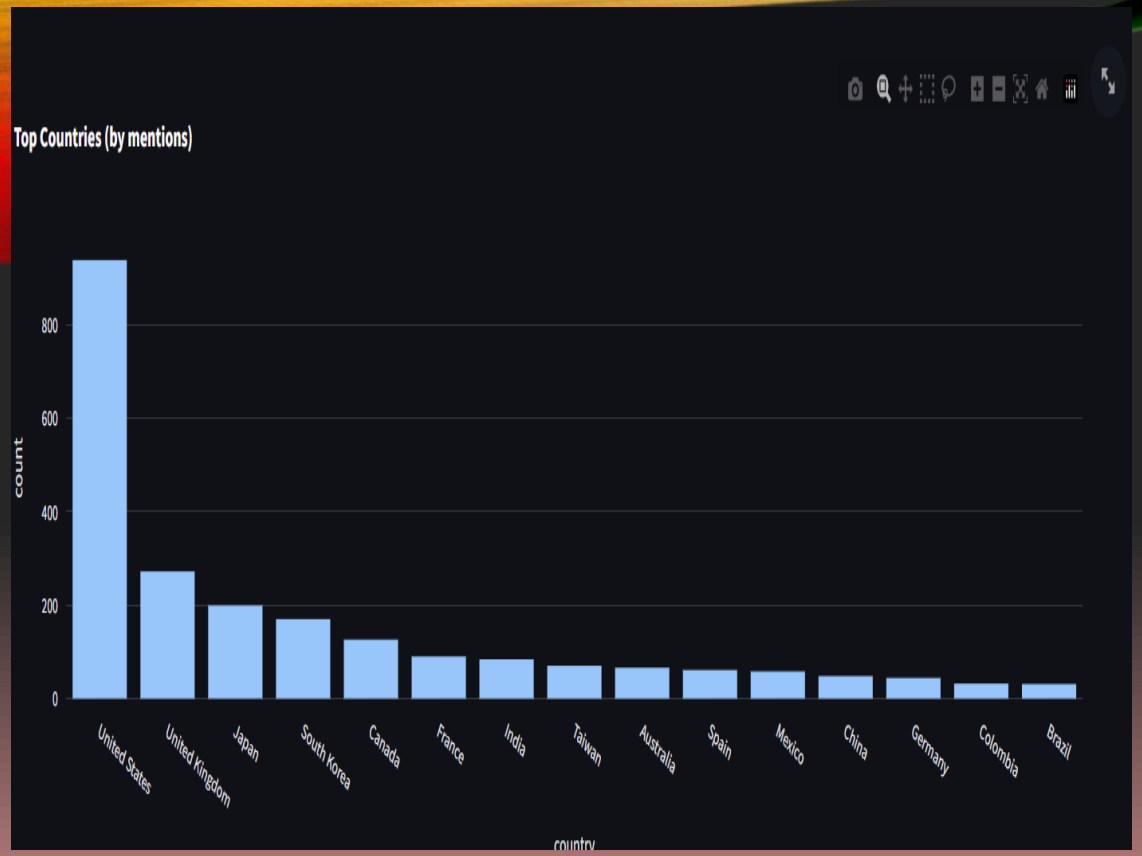
- **Numpy** for data manipulation
- **Matplotlib**
- **Sckit** learn for predictive modelling
- **Joblib** for model serialization
- **Deployment and Interface:**
- **Streamlit** for rapid development
- **Render** for cloud deployment
- **Dataset Source:**
- Kaggle
- Mymoviedb.csv provides rich features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



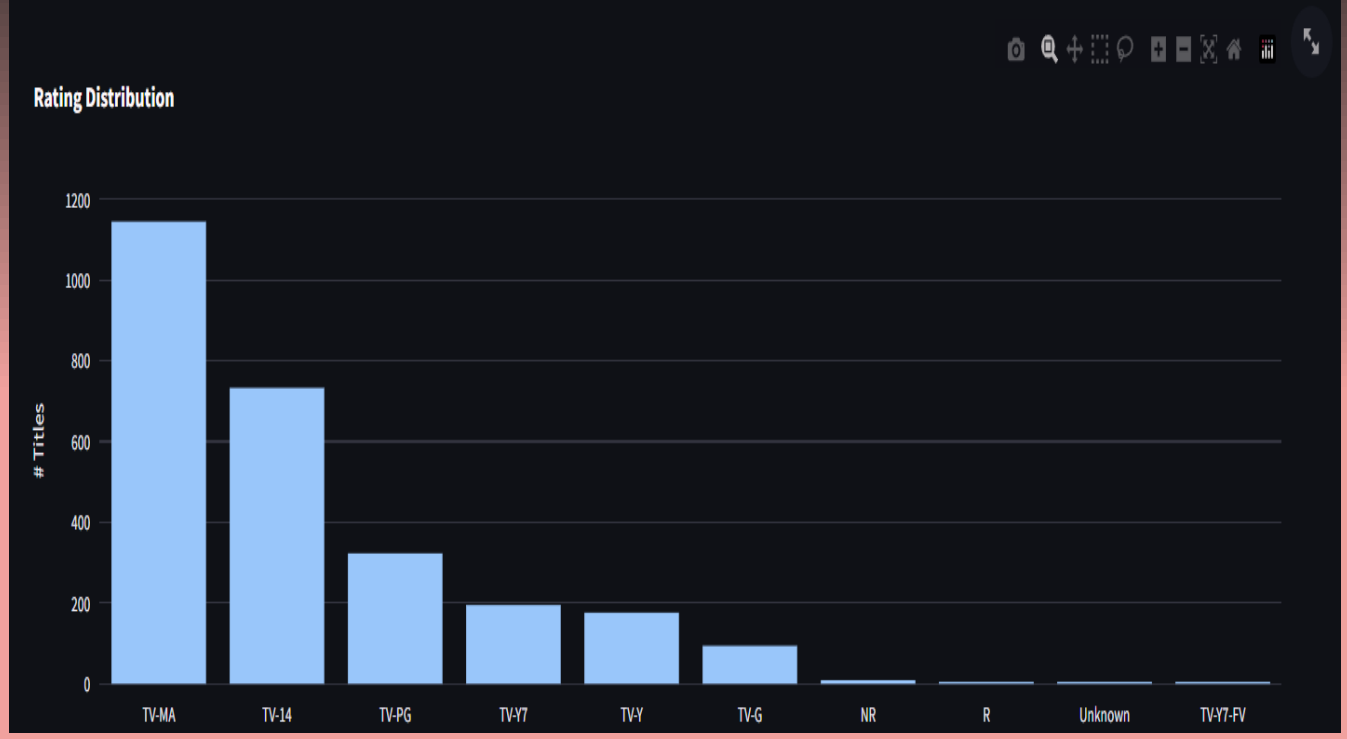
PROJECT WORKING

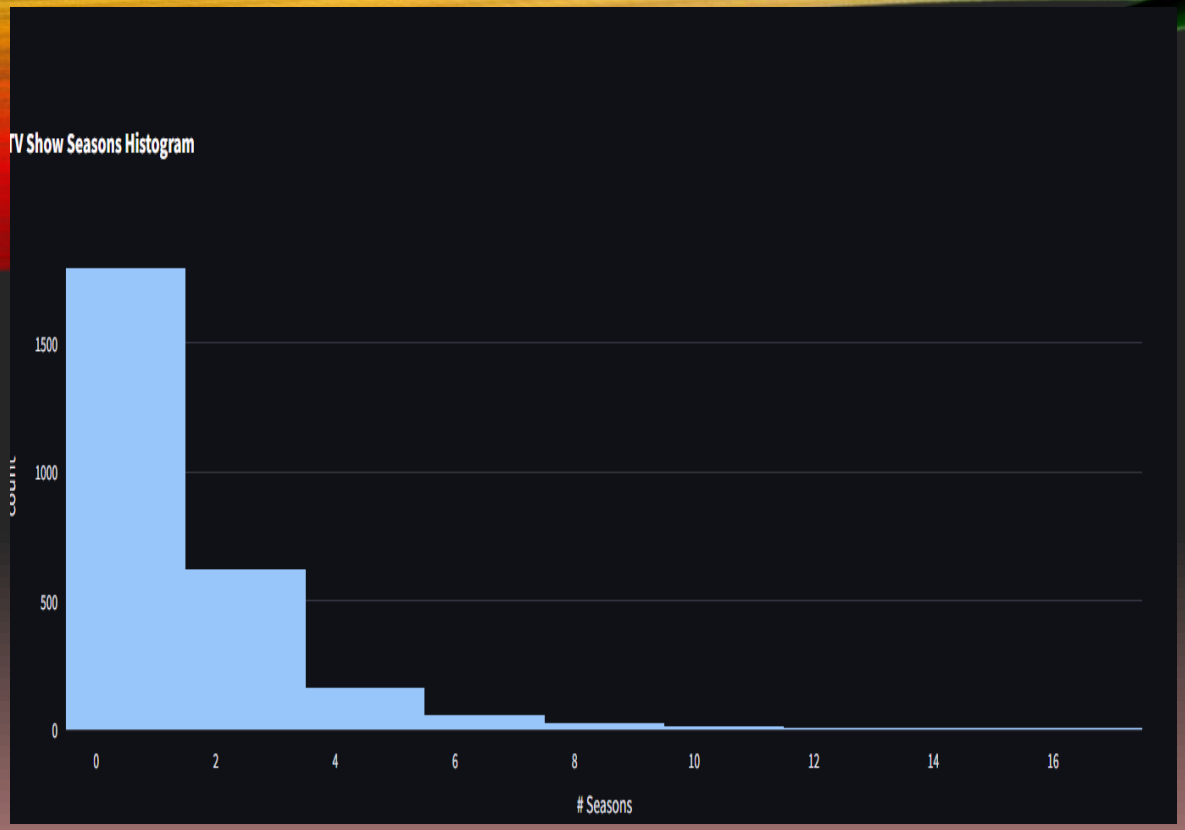


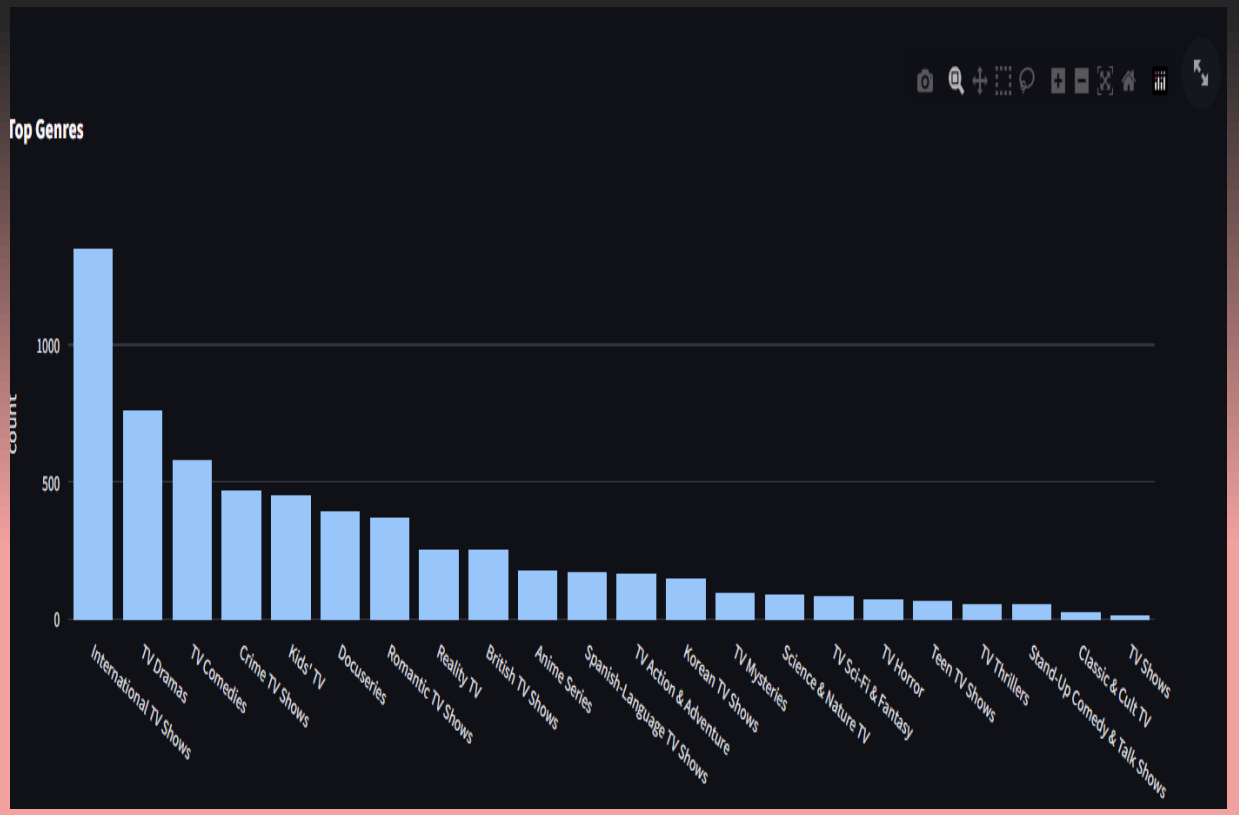


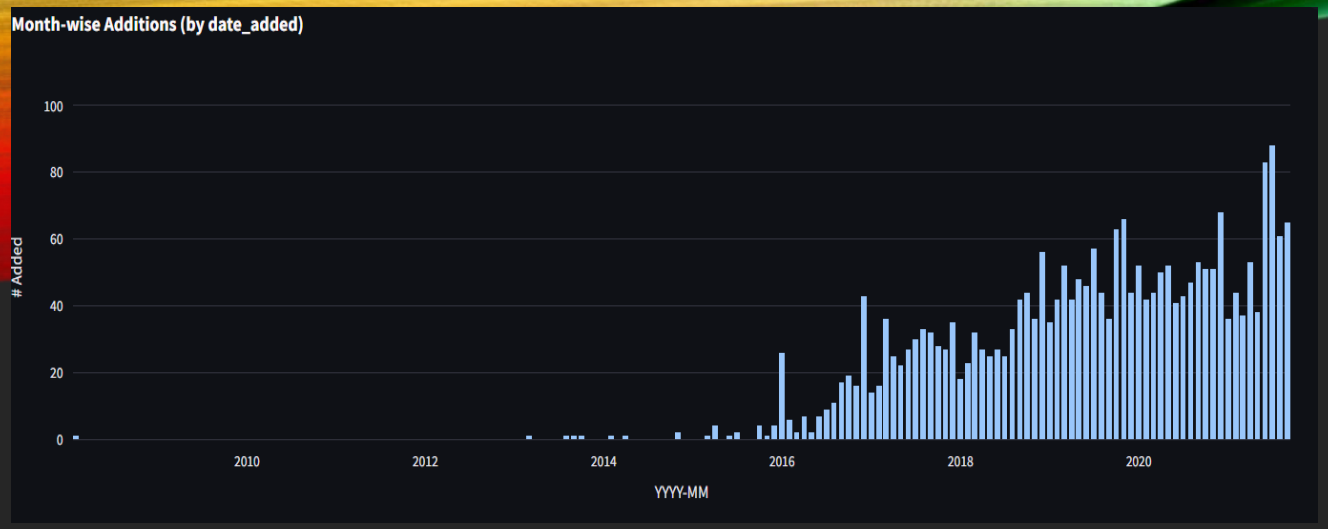


Titles (filtered)	Movies	TV Shows	Unique Countries	Unique Genres
2,676	0	2,676	65	22

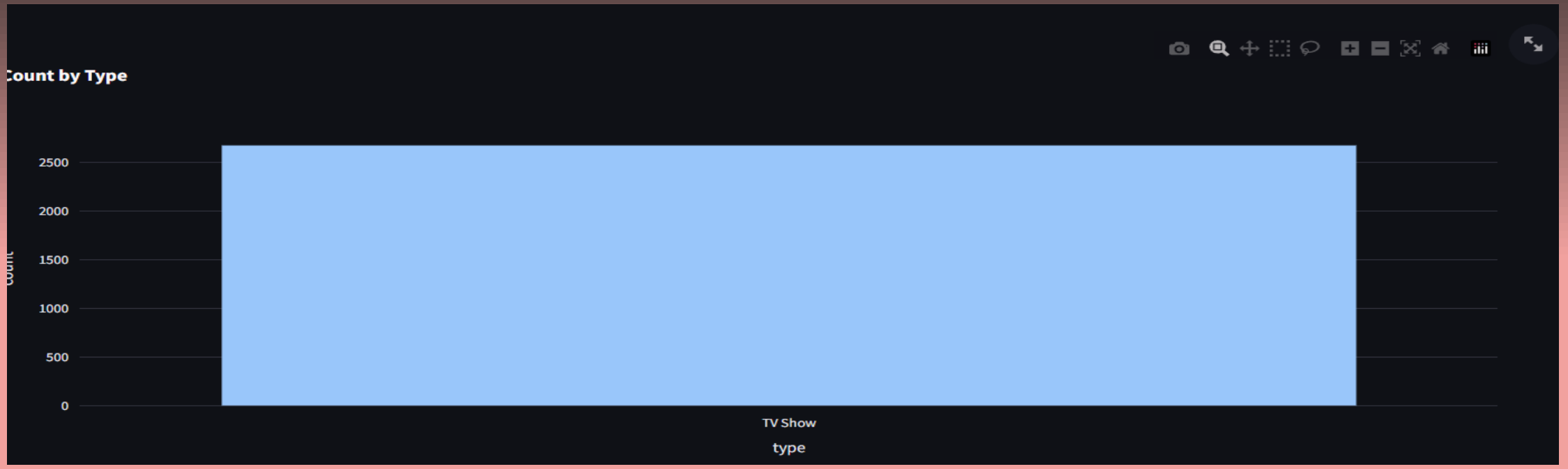



TV Show Seasons Histogram
1500
1000
500
0
0
2
4
6
8
10
12
14
16
Seasons


Top Genres
1000
500
0
International TV Shows
TV Dramas
TV Comedies
Crime TV Shows
Kids' TV
Docuseries
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)
Added
100
80
60
40
20
0
2010
2012
2014
2016
2018
2020
YYYY-MM

Life
World
Love
La
Girl
House
Stories
War
Night
Killer
One
Master
Power
Rangers
Adventure
Friend
Time
Show
King
Serie
City
Crime
Legend
Murder
Secret
Family
Super
Black
Tale
Man
Planet
Last
Day
Happy
Queen
Earth
Good
Dark
Transformers
Marvel
Space
Dragon
Monster
Kingdom
Inside
Back
Big
Song
New
Next
Wild
Game
Rescue
Age
Blood
Little
Bangkok
Call
Great
Force
Story
Money
Boy
Home
History
Journey
Mind
Go
File
Heart
Club
Baby
True
Way
Dead
Ghost
First
Cat
Bear
Detective
American
Dream
Evil
LEGO
Live
Word
Thing
Fate
Squad
Made
Chronicles
Drug
Empire
Power
Road
Kid
Bad
Behind
Beyond
Holiday
Academy
Explained
Ken Burns
Club Friday
Hotel
Christmas
Ninja
Stranger
Hot
Law
Seven
Flower
Classic
Street
Death
Battle
Hunter
Fire
France
Food
Universe
Nature
High
Magic
Lost
Miss
School
Code
Spy
Eye
Heaven
Human
Case
Prince
Mexico
Comedy
Animal
Nailed
Luna
Agent
Town
Mystic
Sex
Deep
Gold
Part
Best
Moon
Cinta
Dear
Rise
Kill
Eat
Dangerous
Oh
Team
Party
Away
Great British
Alien
Castle
White
Film
Women
Online
Truth
Midnight
Return
Pokemon
Young
People
Trial
African
Innocent
Art
Water
Mr
Power
Sin
Year
Iron
Monty Python
Science
Criminal
DreamWorks
Garden
Diaries
Beyond
Best
Part
Sex
Lo
Dad
Cinta
Ninja
Stranger
Christmas
Hotel
Two
Code
Nailed
Luna
Agent
Town
Mystic
Sex
Deep
Gold
Part
Best
Moon
Cinta
Dear
Rise
Kill
Eat
Dangerous
Oh
Team
Party
Away
Great British
Alien
Castle
White
Film
Women
Online
Truth
Midnight
Return
Pokemon
Young
People
Trial
African
Innocent
Art
Water
Mr
Power
Sin
Year
Iron
Monty Python
Science
Criminal
DreamWorks
Garden
Diaries
Beyond
Best
Part


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

APPLICATIONS:

- Content analytics for streaming services
- Personalized recommendation systems based on user trends and genre preferences
- Market research for content acquisition strategy
- Sentiment and pattern analysis for improved user engagement and usage statistics

PROBLEMS FACED & SOLUTIONS:

Challenges

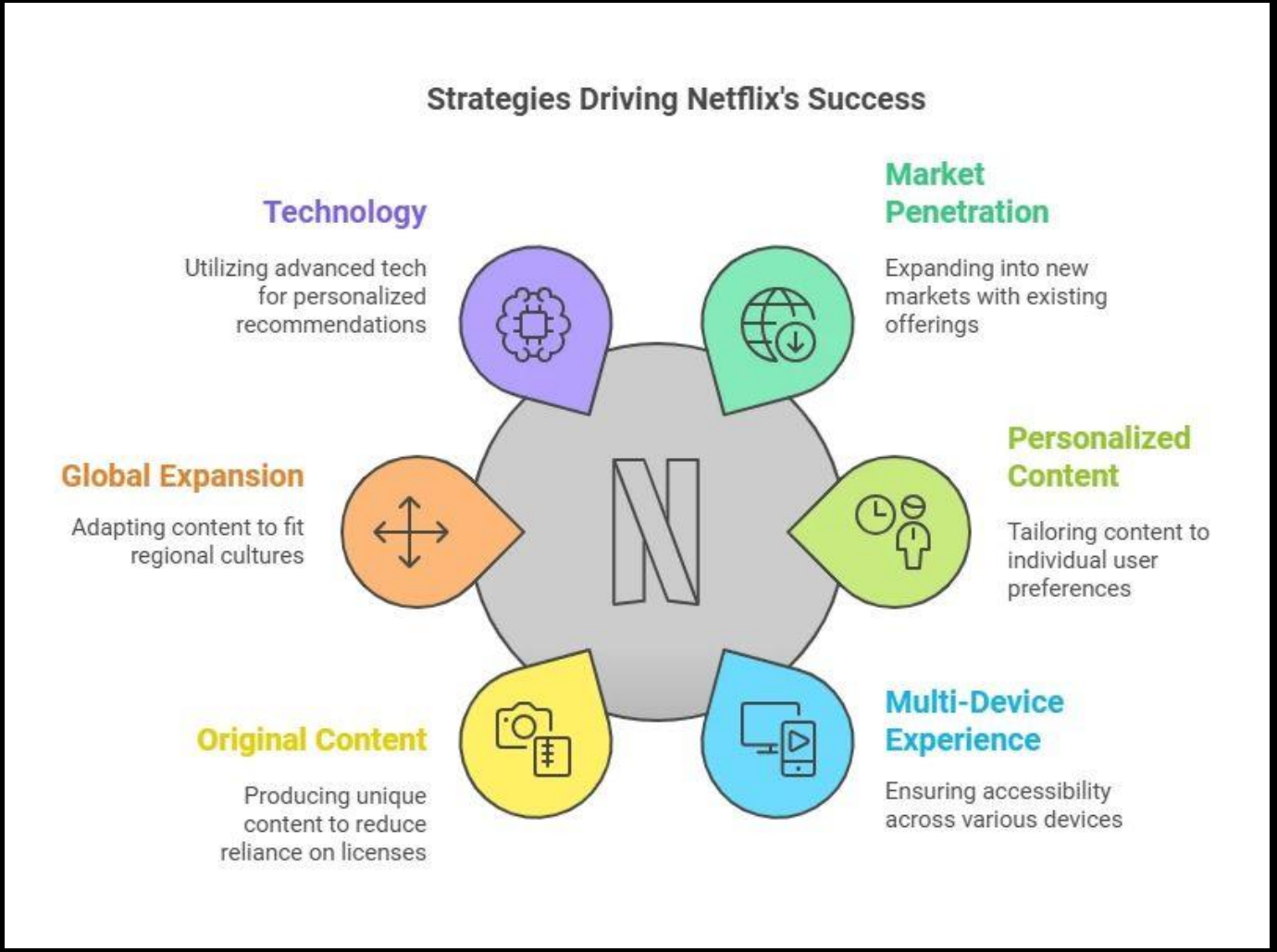
- Some data columns were in incorrect or mixed formats (like dates and time durations).
- Needed a way to make the project more engaging and user-friendly.
- It was hard to decide which visualizations best represent the data.
- Large dataset with many attributes

Solutions

- Reformatted those columns into the right data types to make analysis smoother.
- Developed an interactive Streamlit dashboard where users can explore data by selecting different options.
- Tried different chart types like bar graphs, pie charts, and word clouds, then chose the ones that made the insights clear and easy to understand.
- Preprocessing and efficient filtering using Pandas and Scikit-learn

FUTURE SCOPE:

- Use machine learning to predict popular genres and content types
- Advanced clustering and recommendation systems with deep learning
- Integrate user interaction data for personalized content suggestions
- Expand analysis with real-time streaming data and sentiment analysis



THANK YOU