EDA ON NETFLIX & TV SHOWS

PRESENTED BY : - SOUMYA DAS
GUIDE:- MISS NAINA DEVI
TECHNO INDIA UNIVERSITY | B.TECH CSE 3RD YEAR

🎯 AGENDA

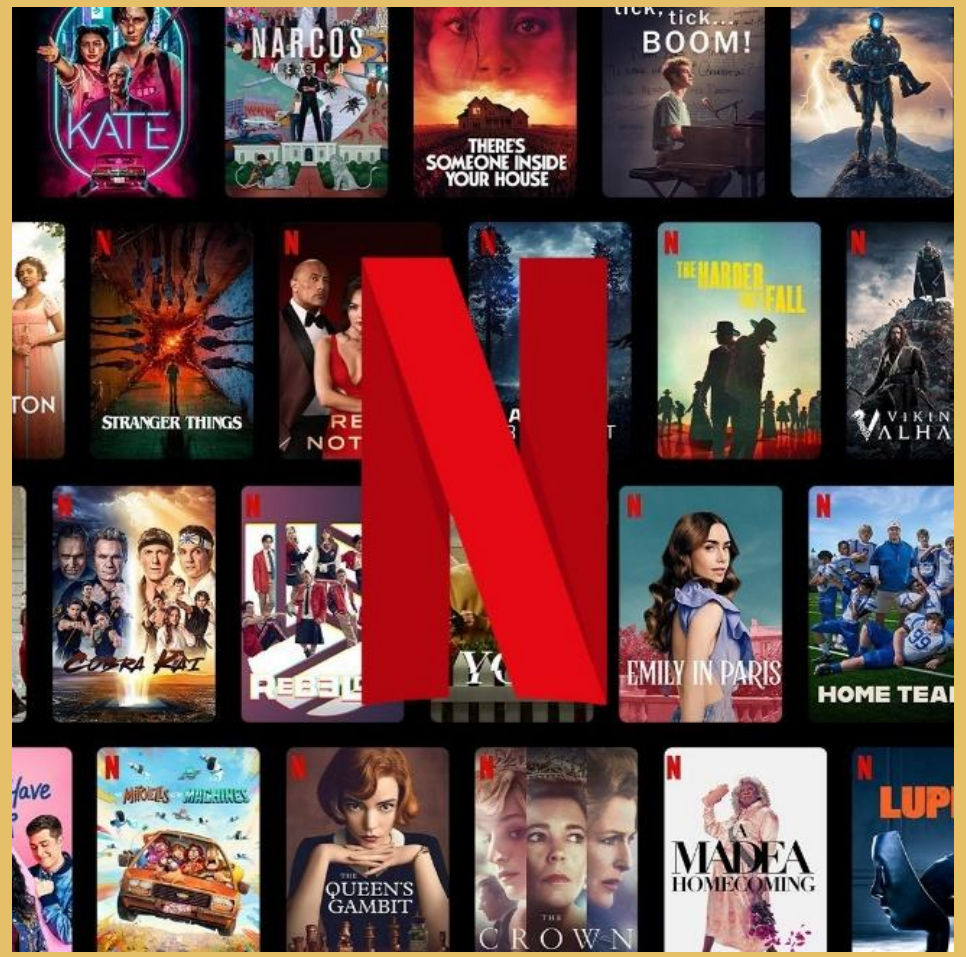
This project explores the Netflix dataset, which contains details about movies available on the platform. The objective is to perform exploratory data analysis (EDA) to uncover insights about genres, ratings, release dates and years etc.

INTRODUCTION

Netflix was founded in 1997 by Reed Hastings and Marc Randolph. The company started as a DVD-rental service, but it quickly transitioned to streaming in 2007. Netflix has grown rapidly over the past few decades, and it is now one of the most popular streaming services in the world. It is one of the most popular media and video streaming platforms

This project explores the Netflix dataset, which contains details about movies available on the platform

This project analyzes extensive data of Netflix shows and an EDA is being prepared in the Back-end which is done with the help of Jupyter Notebook ,also requires anaconda navigator and prompt.
The frontend is being prepared by Streamlit and then deployed with the help of Render which is basically used for cloud deployment and accessibility.



TECHNICAL ECOSYSTEM : TOOLS & TECHNOLOGIES

We employed a robust set of tools for data handling, model training, and deployment, ensuring a scalable and efficient solution.

CORE LIBRARIES

Pandas , **Numpy** for data manipulation , **Matplotlib** , **Seaborn** for visualization ; **Sckit-learn** for predictive modelling ; **Joblib** for model serialization

DEPLOYMENT AND INTERFACE

Streamlit for rapid development of the interactive web application; **Render** for cloud deployment and accessibility

DATA AND ENVIRONMENT

DATASET: mymoviedb.csv provides rich featues for **Netfix** shows EDA.Development in **Jupyter Notebook** and version control with **GitHub**

PROJECT WORKING

Dashboard Menu

Click on any section to visualize data:

Select Graph:

- 1 - Dataset Overview
- 2 - Genre Distribution
- 3 - Release Date Column Distribution
- 4 - Top 10 Genres
- 5 - Popularity vs Vote Count
- 6 - Popularity Share by Genre (Pie Chart)
- 7 - Highest & Lowest Popularity Movies
- 8 - Correlation Heatmap

Netflix Movies & TV Shows EDA Dashboard

Dataset Overview

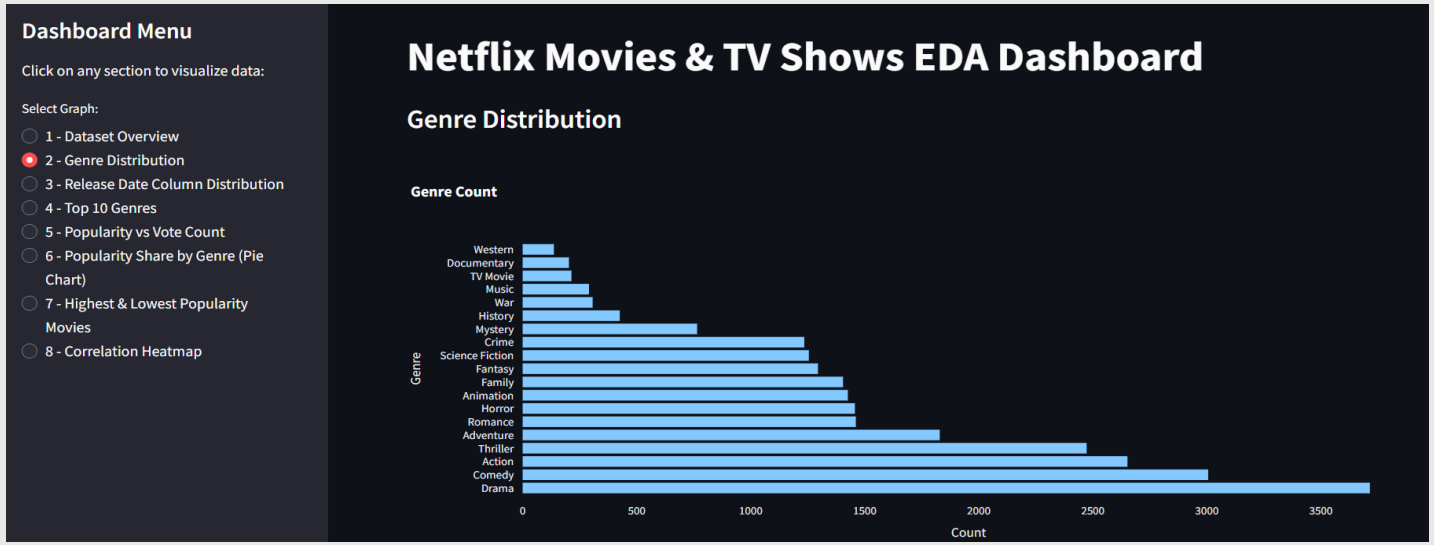
Shape: (25552, 7)

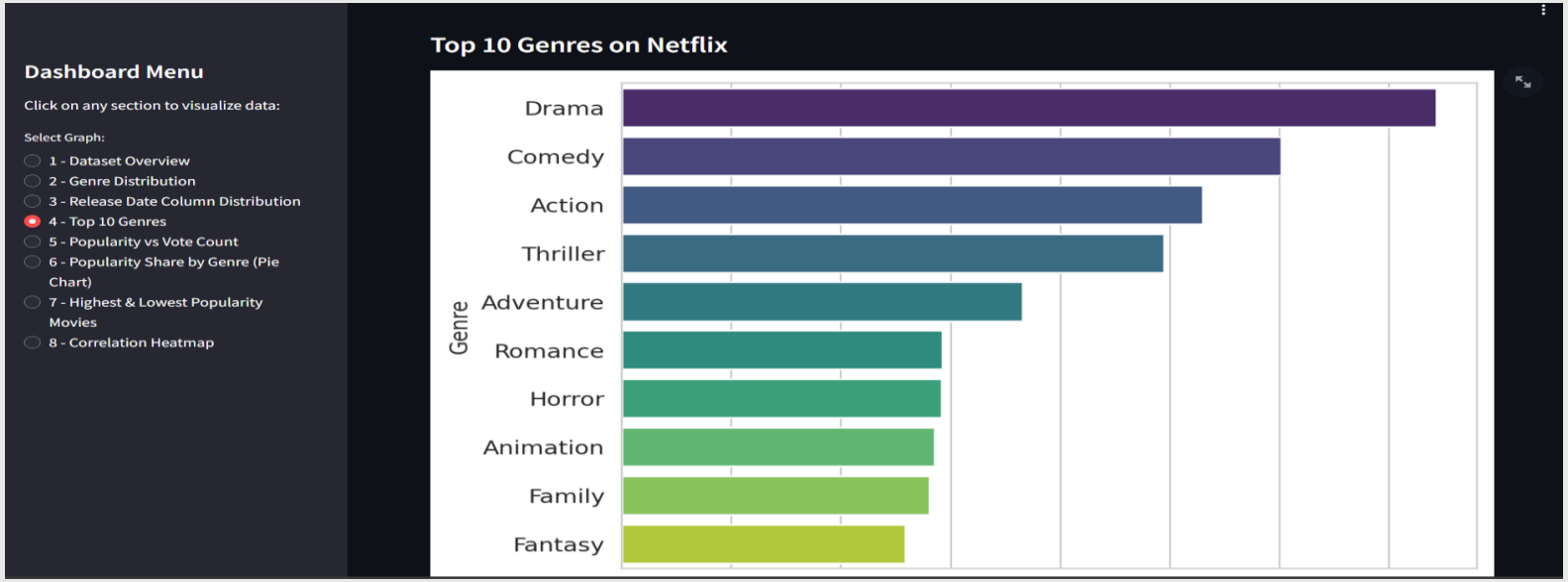
	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre	Year
0	1,970	Spider-Man: No Way Home	5,083.954	8,940	popular	Action	1,970
1	1,970	Spider-Man: No Way Home	5,083.954	8,940	popular	Adventure	1,970
2	1,970	Spider-Man: No Way Home	5,083.954	8,940	popular	Science Fiction	1,970
3	1,970	The Batman	3,827.658	1,151	popular	Crime	1,970
4	1,970	The Batman	3,827.658	1,151	popular	Mystery	1,970

Columns:


```
[
  0 : "Release_Date"
  1 : "Title"
  2 : "Popularity"
  3 : "Vote_Count"
  4 : "Vote_Average"
  5 : "Genre"
```

PROJECT WORKING





PROJECT WORKING

Dashboard Menu

Click on any section to visualize data:

Select Graph:

- 1 - Dataset Overview
- 2 - Genre Distribution
- 3 - Release Date Column Distribution
- 4 - Top 10 Genres
- 5 - Popularity vs Vote Count
- 6 - Popularity Share by Genre (Pie Chart)
- 7 - Highest & Lowest Popularity Movies
- 8 - Correlation Heatmap

Netflix Movies & TV Shows EDA Dashboard

Highest & Lowest Popularity Movies

Most Popular Movie(s):

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre	Year
0	1,970	Spider-Man: No Way Home	5,083.954	8,940	popular	Action	1,970
1	1,970	Spider-Man: No Way Home	5,083.954	8,940	popular	Adventure	1,970
2	1,970	Spider-Man: No Way Home	5,083.954	8,940	popular	Science Fiction	1,970

Least Popular Movie(s):

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre	Year
25,546	1,970	The United States vs. Billie Holiday	13.354	152	average	Music	1,970
25,547	1,970	The United States vs. Billie Holiday	13.354	152	average	Drama	1,970
25,548	1,970	The United States vs. Billie Holiday	13.354	152	average	History	1,970
25,549	1,970	Threads	13.354	186	popular	War	1,970
25,550	1,970	Threads	13.354	186	popular	Drama	1,970
25,551	1,970	Threads	13.354	186	popular	Science Fiction	1,970

🌎 REAL LIFE APPLICATIONS AND VALUE

REAL-LIFE APPLICATIONS & VALUE:

🎬 **CONTENT INSIGHTS**: HELPS STREAMING PLATFORMS IDENTIFY POPULAR GENRES, AUDIENCE PREFERENCES, AND CONTENT TRENDS

📈 **DATA-DRIVEN DECISIONS**: ENABLES PRODUCERS AND MARKETERS TO PLAN RELEASES AND PROMOTIONS BASED ON REAL DATA.

🤖 **FOUNDATION FOR RECOMMENDATIONS**: FORMS THE BASE FOR MACHINE LEARNING MODELS THAT SUGGEST MOVIES OR SHOWS TO USERS.

💡 **INDUSTRY RELEVANCE**: REFLECTS REAL-WORLD DATA ANALYTICS USED BY NETFLIX, AMAZON PRIME & DISNEY+.

🧠 **SKILL DEVELOPMENT**: STRENGTHENS PYTHON, DATA VISUALIZATION, AND ANALYTICAL THINKING — KEY SKILLS FOR DATA SCIENCE CAREERS.

📊 **BUSINESS VALUE**: CONVERTS RAW ENTERTAINMENT DATA INTO ACTIONABLE INSIGHTS THAT IMPROVE USER ENGAGEMENT AND RETENTION.

CHALLENGES ENCOUNTERED AND SOLUTIONS

IN BACK-END

CHALLENGES:

- **Data Cleaning Complexity** – Missing values, inconsistent genres, and date formats.
- **Large Dataset Handling**
- **Choosing the right visuals**
- **Maintaining Data Accuracy**

SOLUTIONS

- Used pandas functions like dropna(), astype(), and to_datetime() to clean and standardize data.
- Optimized data loading with lineterminator='\n' and selective column dropping to improve speed.
- Tested multiple plot types (bar, scatter, pie, line) using Seaborn, Matplotlib & Plotly for best data clarity.
- Verified data after each transformation and saved a cleaned CSV for consistent use.

CHALLENGES ENCOUNTERED AND SOLUTIONS

IN FRONT-END

CHALLENGES:

- Integration with Backend
- Graph Display Errors
- UI Layout & Navigation
- Deployment Issues (App.py case sensitivity)

SOLUTIONS

- Imported backend (EDA) functions into Streamlit app and used modular structure to link data and visuals.
- Adjusted figure sizes and used Streamlit's built-in plotting functions st.pyplot(), st.plotly_chart().
- Used Streamlit sidebar for menu options and created sections for each visualization category.
- Renamed files properly and followed Streamlit deployment guidelines for Render/Cloud platforms.

FUTURE SCOPE

TO ELEVATE THE PROJECT INTO A COMPREHENSIVE,PRODUCTION – READY SOLUTION,WE PLANNED TO IMPLEMENT SEVERAL ADVANCED FEATURES

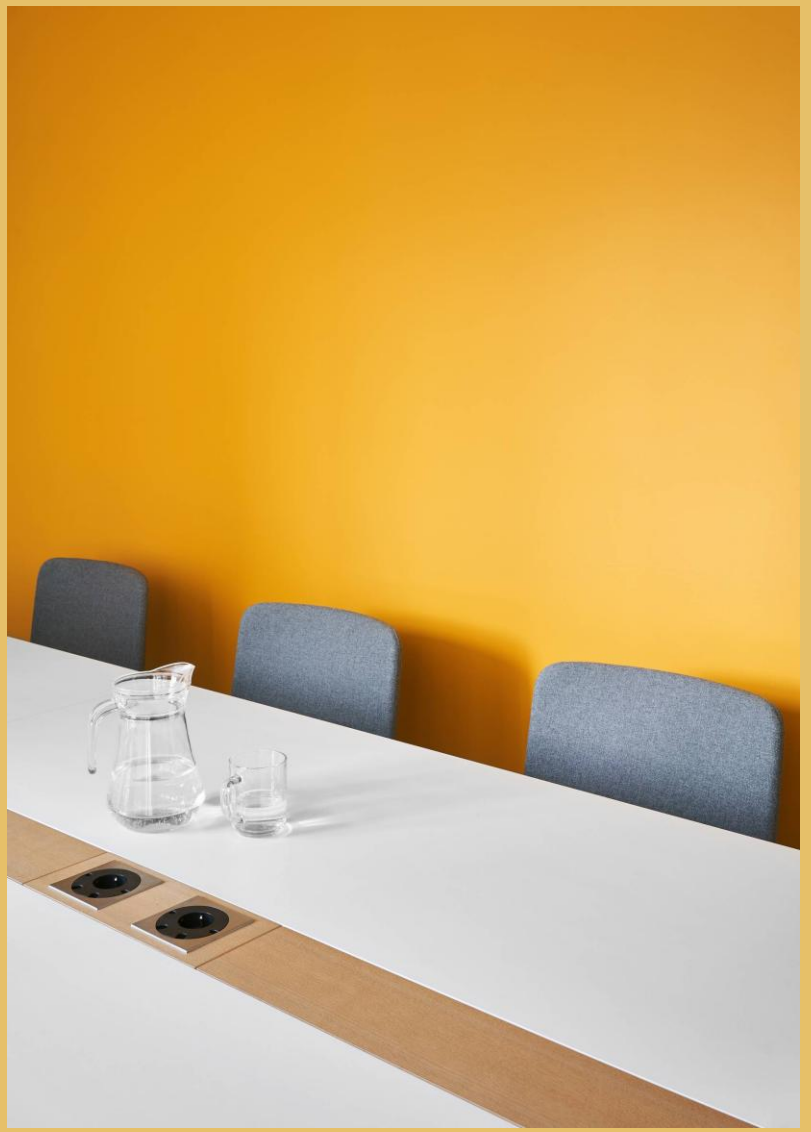
- **Content Recommendation System:**
  Use ML algorithms (like Collaborative Filtering or Content-Based Filtering) to suggest movies/shows to users based on preferences.
- **Predictive Analytics:**
  Build models to predict movie popularity, ratings, or success probability based on factors like genre, cast, or release year.
- **Sentiment Analysis:**
  Integrate external reviews or comments data to analyze audience sentiment using Natural Language Processing (NLP).
- **Interactive Dashboard Upgrade:**
  Enhance the Streamlit dashboard with real-time ML predictions, search filters, and recommendation previews.

PROJECT LINK

ACCESS THE LIVE DEPLOYMENT:

THE COMPLETE INTERACTIVE DASHBOARD IS DEPLOYED AND ACCESSIBLE VIA RENDER:

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



THANK YOU