Data Insights on

NETFLIX

This project aims to uncover patterns in Netflix's vast library of films and TV shows, exploring how content types have evolved over time and identifying popular genres across different regions.

Play

More Info

TEAM – BLACK HAWKS

Rithwik Vamshi
rathin3@uic.edu
GIT: rithwikvamshi

Rohith Kumar
vchenn6@uic.edu
GIT: Rohith1110

Bhargavaram
bmupp@uic.edu
GIT: BhargavaramMuppuru

Asritha
apidik2@uic.edu
GIT :asrithap12

Sanjna
schipp3@uic.edu
GIT : sanjna19

GIT REPOSITORY

INTRODUCTION

Scope: Analyze over 8,000 movies and TV shows

Primary Dataset: In-depth data across 12 categories.

Extended Analysis: Incorporate external ratings from IMDB & Rotten Tomatoes to enrich our understanding.

Objective: Provide a comprehensive overview of Netflix's offerings and viewer reception, enhancing the strategic insights into streaming content trends.

PROBLEM STATEMENTS

1. How the popularity & evolution of different content types changed over time?
2. What are the current content strategies for various countries?
3. Is it possible to predict the success of new titles?
4. Who are the key industry figures that Netflix should consider partnering with?

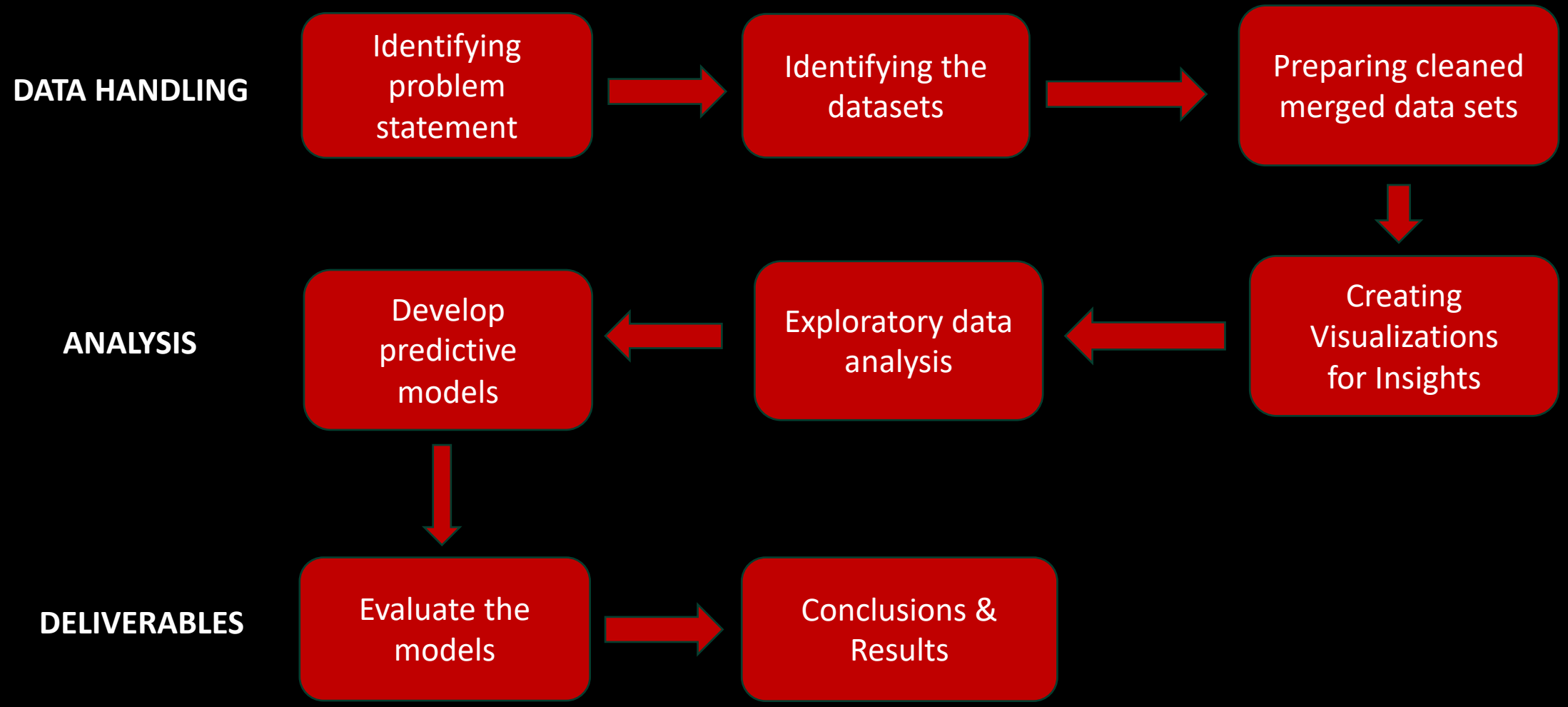
DATASET

1. **Sources**: Merged dataset from Kaggle's Netflix titles and IMDb Ratings.
2. **Integration Effort**: Datasets have been combined using title matching, creating a rich dataset for in-depth analysis.
3. **Dataset Characteristics**:

- Total Entries: 6094 combined records.
- Types of Data:
  - Categorical: Type (Movie/TV Show), country, content rating, genre.
  - Numerical: Release year, various rating scores.
  - Textual: Title, description.
- Features: Comprehensive metadata from Netflix, critical ratings from external source, content availability.

4. **Usage**: Content performance analysis & viewer preferences study.

WORKFLOW


DATA HANDLING
Identifying problem statement
Identifying the datasets
Preparing cleaned merged data sets
ANALYSIS
Creating Visualizations for Insights
Exploratory data analysis
Develop predictive models
DELIVERABLES
Evaluate the models
Conclusions & Results

OUR APPROACH

1

Descriptive analysis

Summarize current content trends

2

Predictive analysis

Forecast viewer preferences

3

Prescriptive analysis

Guide in strategic decision-making

OUR APPROACH

- We used Python to analyze and visualize data
- Before diving into our analysis of Netflix data, we had to make sure the data was in good shape This involved a few important steps such as Data cleaning
- As our dataset exclusively contained categorical data, we incorporated an external dataset to facilitate further analysis
- Now, we'll begin by using basic statistics to analyze and visualize Netflix data

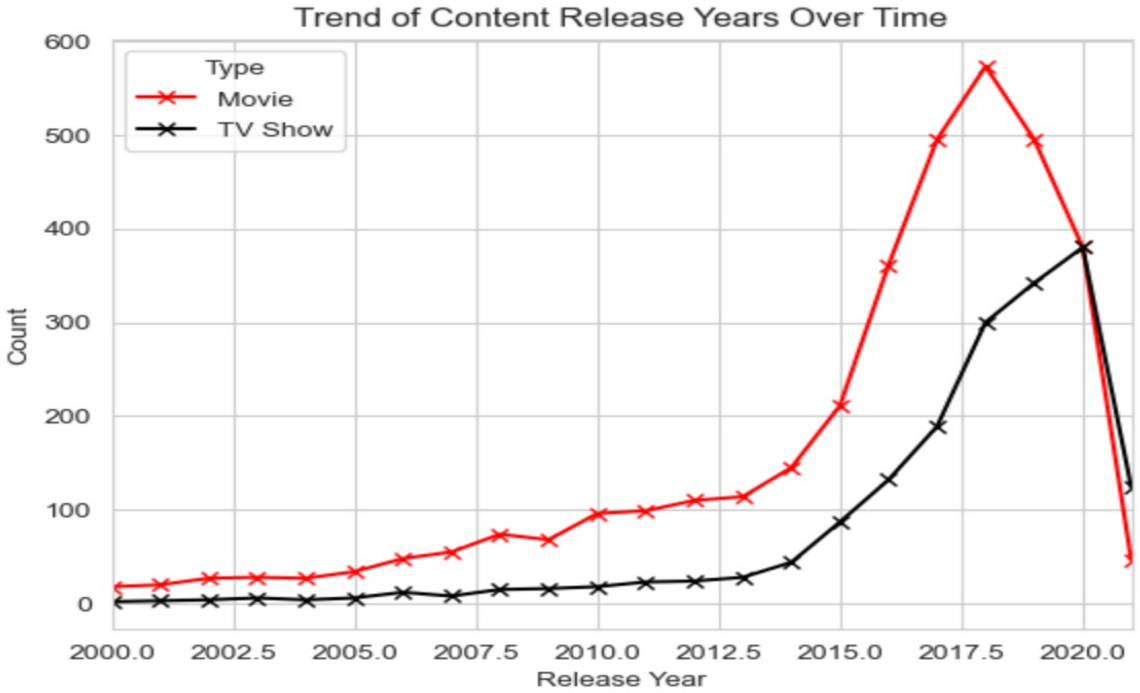
Descriptive Analytics

How the popularity & evolution of different content types changed over time?

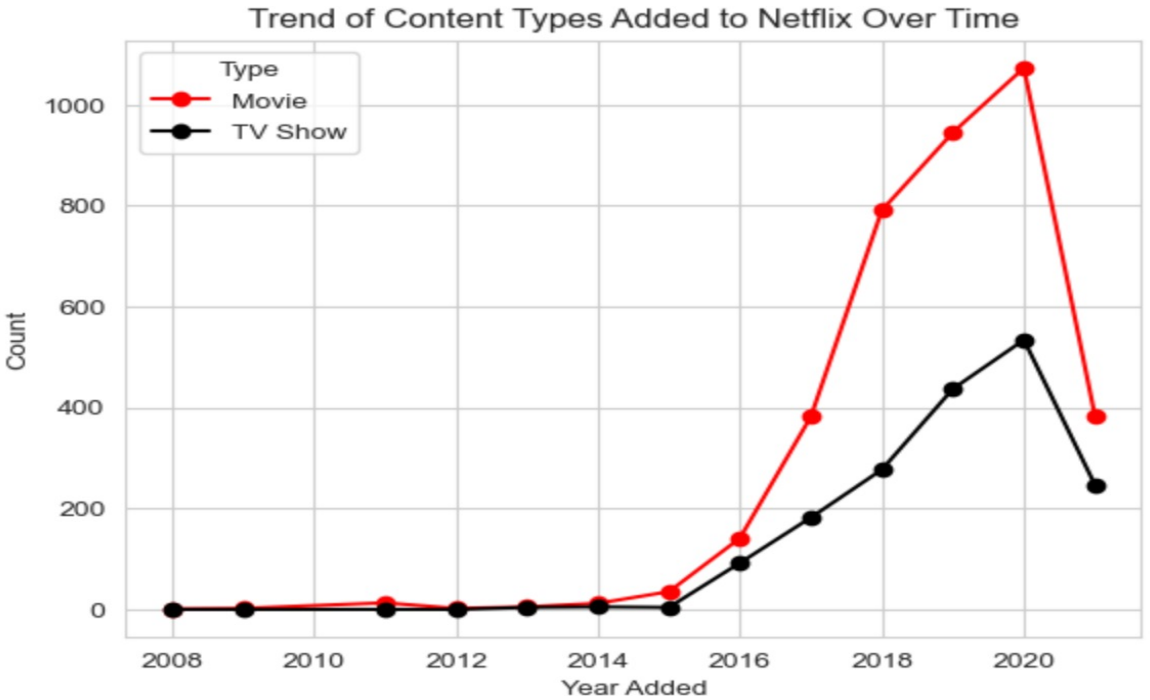
Using time-series analysis of Netflix content types (Movies and TV Shows) , we examine their changing popularity by considering the date_added and release_year fields

Our analysis will focus on tracking trends over the years, assessing the evolving proportion of Movies versus TV Shows, and studying how the release years of content relate to their addition to Netflix

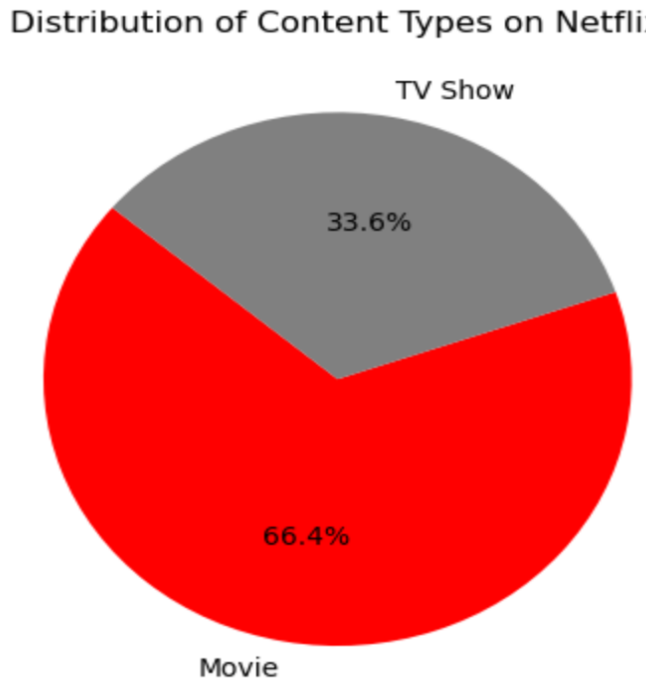
By exploring these aspects, we aim to gain insights into the historical dynamics of content types on Netflix



Netflix's content strategy since the year 2000 emphasizes a substantial inclusion of recent materials, showcasing the platform's commitment to maintaining an updated and contemporary library.



The data indicates an overall increase in Netflix's library, with a particularly sharp rise in recent years. Movies are being added more than TV shows, highlighting the platform's expansive content strategy.



Content Type: The pie chart reveals that on Netflix, Movies constitute a larger share of content than TV Shows.

Conclusion

Netflix's catalog leans more towards movies, although TV Shows have experienced considerable growth recently. The addition of newer releases indicates a subscriber demand for fresh content.

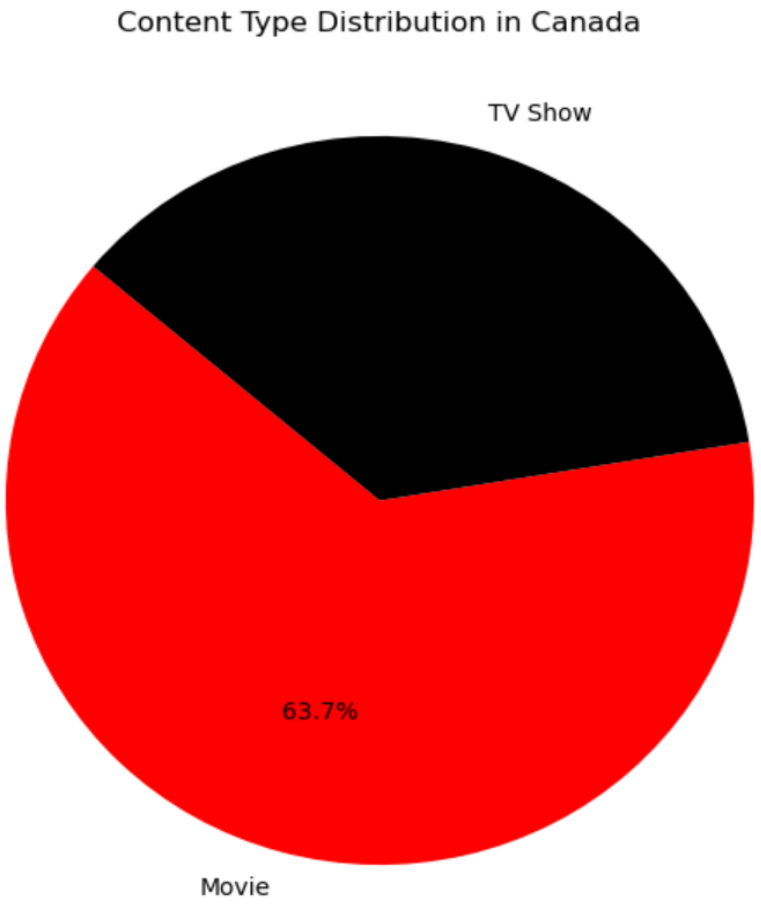
Part 2:

What current content strategies can be seen for different countries?

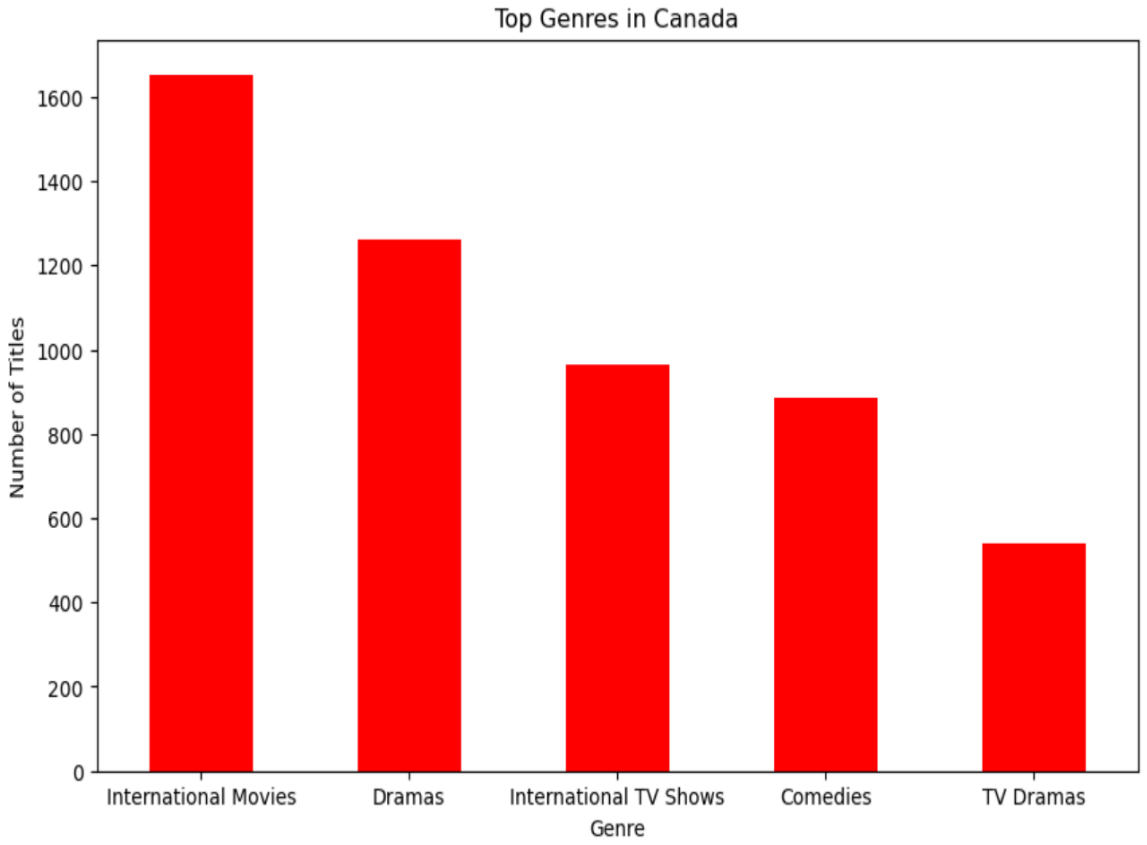
To answer this question, we started by analyzing the "Country Availability" column, which included counts of combinations of countries where the same titles were available.

We then proceeded to examine the prevalent content genres, types, languages of the content and ratings in each country.

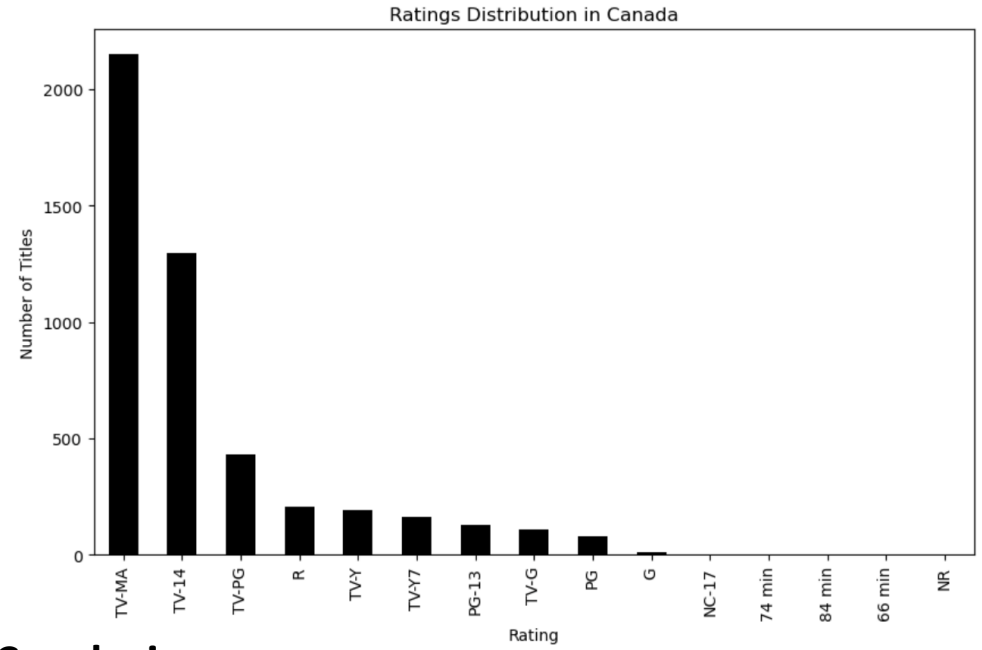
Let's look into Canada’s content strategies.



Movies: 3,041 titles
TV Shows: 1,734 titles

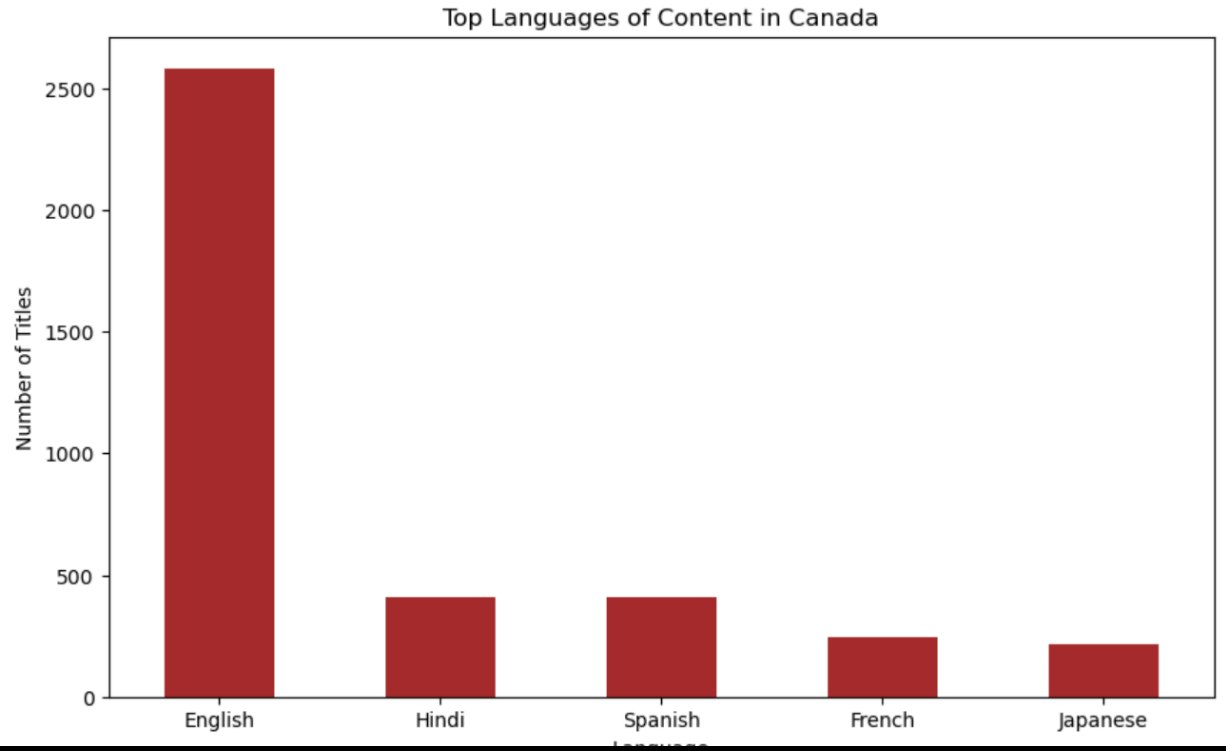


International Movies: 1,653 titles, **Dramas**: 1,263 titles, **International TV Shows**: 966 titles, **Comedies**: 885 titles, **TV Dramas**: 541 titles



Ratings Distribution: TV-MA (Mature Audience): 2,149 titles, TV-14 (Parents Strongly Cautioned): 1,299 titles, TV-PG (Parental Guidance Suggested): 430 titles, R (Restricted): 205 titles

Primary Languages: English: 2,582 titles, Hindi: 410 titles, Spanish: 408 titles, French: 247 titles, Japanese: 215 titles



Conclusion:

The Canadian market has a strong presence of movies, especially international movies and dramas. There's also a significant amount of content rated for mature audiences

The prevalence of Hindi and Spanish titles reflects a diverse linguistic audience

Predictive Analytics

Can we predict which new titles will be hits?

IMDb score is used as a proxy for hit status. We used EDA that involves identifying potential success indicators through feature selection.

For this, with the preprocessed data, we trained the model to perform regression. The model analyzes historical data to find patterns that correlate with high viewership, critical acclaim, or other measures of success.

To assess model performance, relevant metrics such as Mean Squared Error (MSE) and R-squared for regression tasks are employed.

The RandomForestRegressor model has been trained and tested on the data
The regression report shows the following metrics:

R- square	0.8687244053500376
Mean Squared Error (MSE)	0.1776949199255121

Conclusion:

The model's performance is quite strong, which is promising for using it to predict whether new titles will be hits based on the defined IMDb score threshold. We can determine the hit status of a title if the predicted IMDb score is above or below a certain limit.

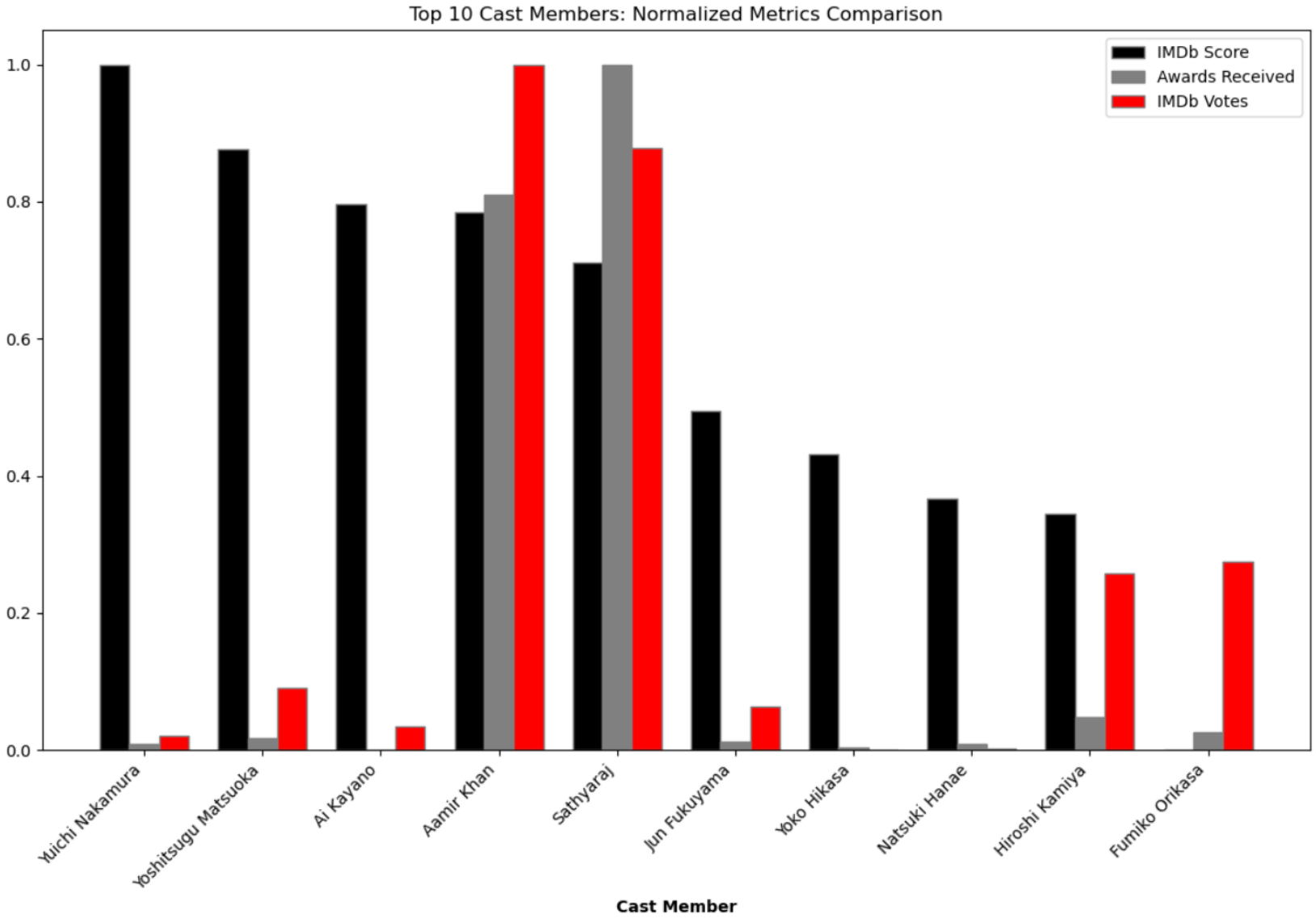
```
Please enter the details for the prediction:
Enter type (Movie/TV Show): Movie
Enter director: Steven Spielberg
Enter country: United States
Enter genre: Fantasy
Predicted IMDb Score: 6.19
```

Prescriptive Analytics

Which industry figures should Netflix partner with for new content?

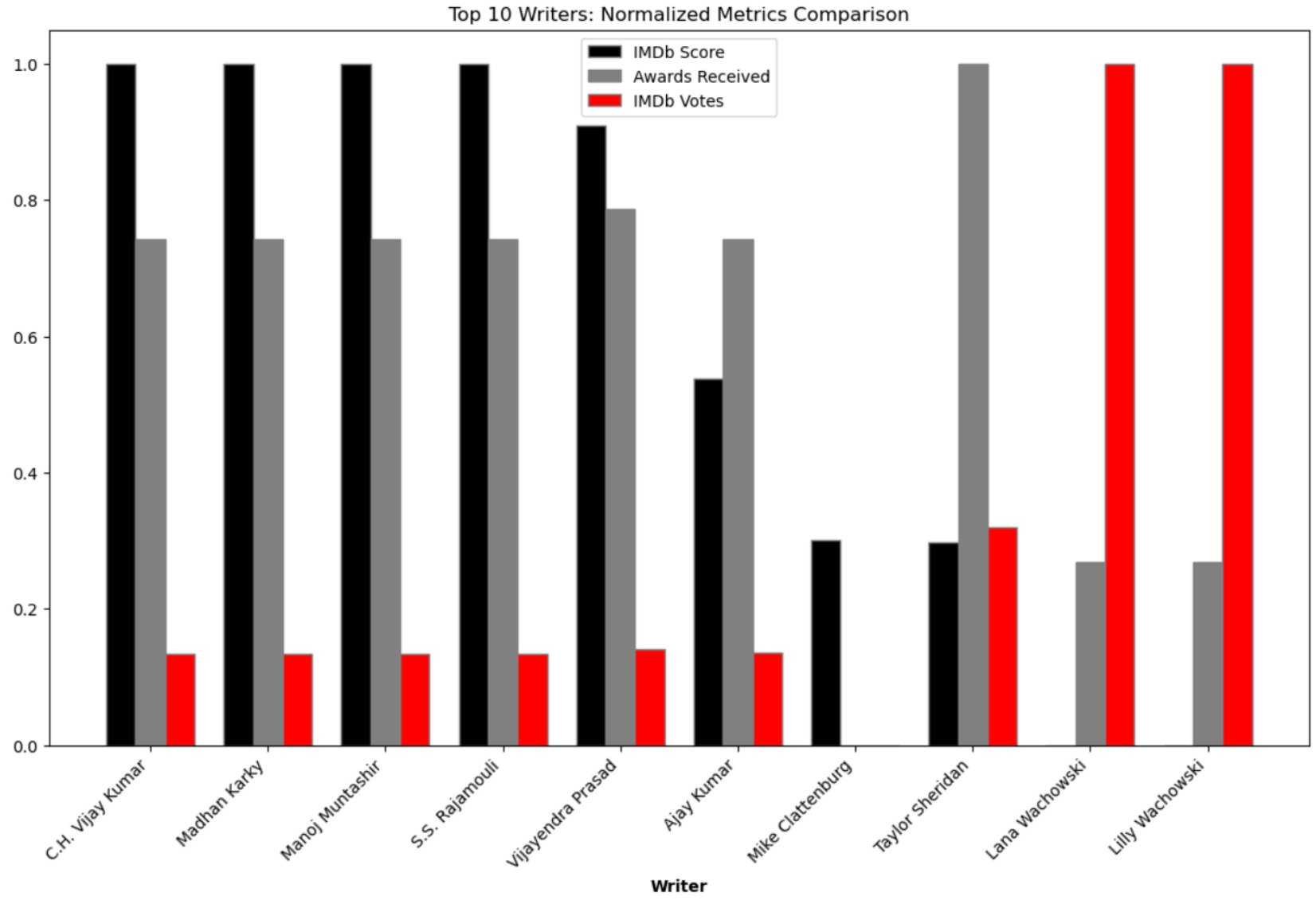
To answer the question, we will look at various factors such as:

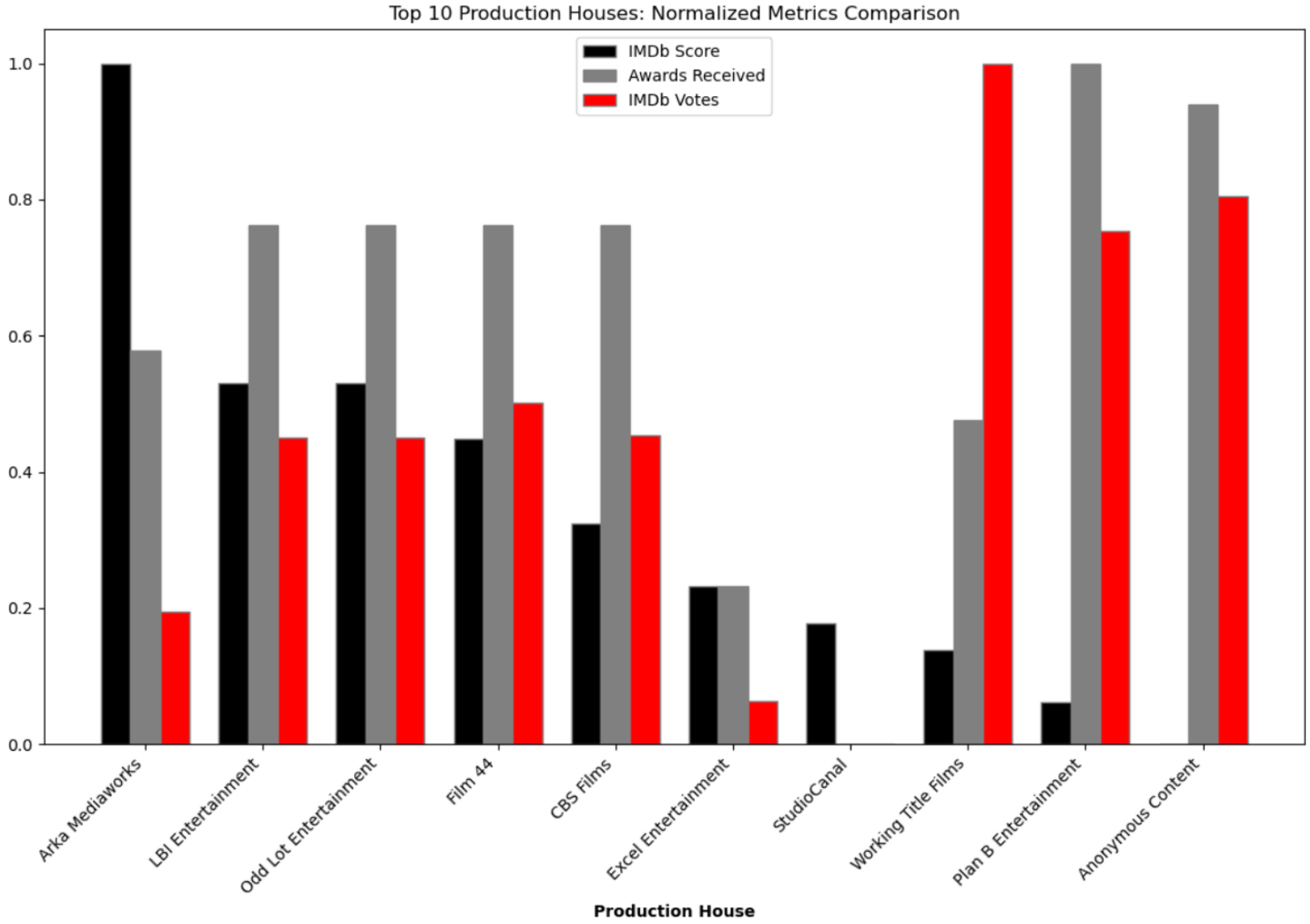
- Directors with the highest number of highly-rated films
- Actors who frequently appear in highly-rated or popular films
- Writers associated with successful films
- Production houses with a strong track record of successful films

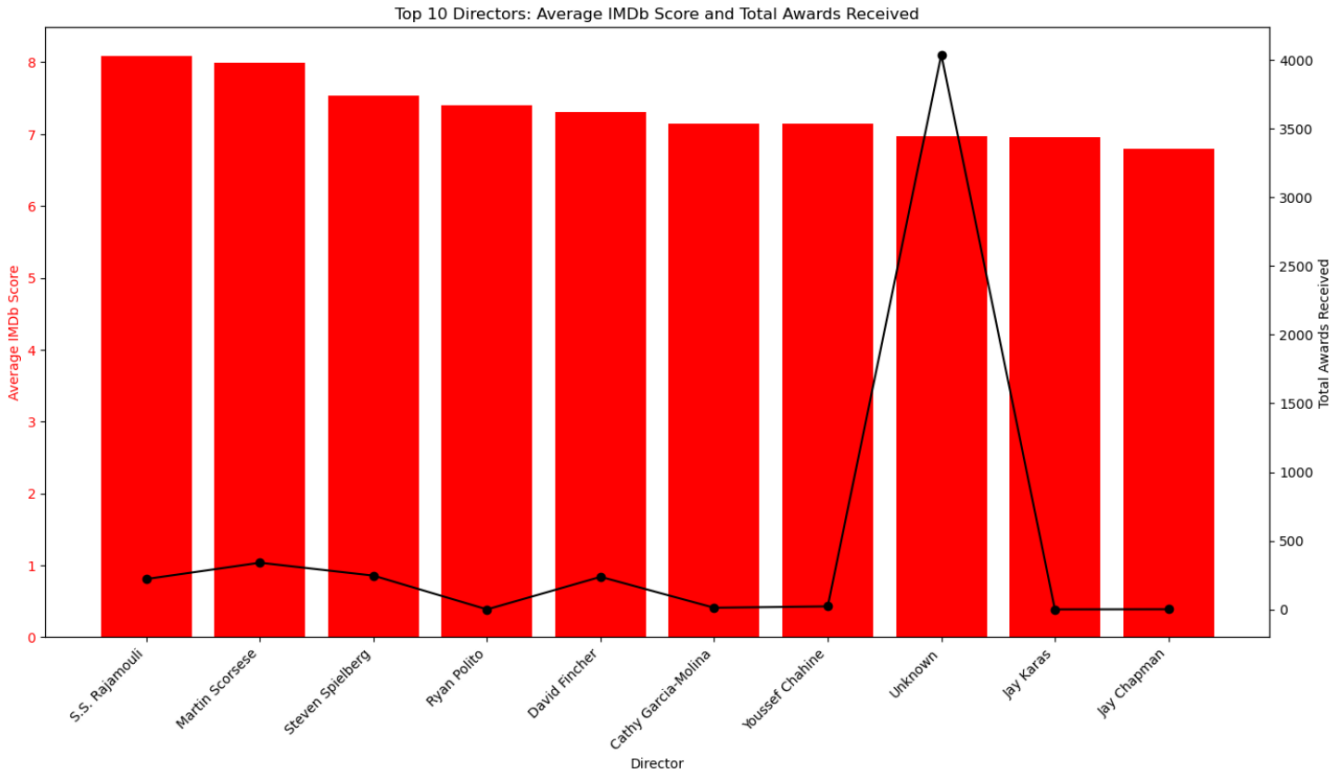


48:35

Prescriptive analysis Part 1


Top 10 Writers: Normalized Metrics Comparison
IMDb Score
Awards Received
IMDb Votes
1.0
0.8
0.6
0.4
0.2
0.0
C.H. Vijay Kumar
Madhan Karky
Manoj Muntashir
S.S. Rajamouli
Vijayendra Prasad
Ajay Kumar
Mike Clattenburg
Taylor Sheridan
Lana Wachowski
Lilly Wachowski
Writer


Top 10 Production Houses: Normalized Metrics Comparison
IMDb Score
Awards Received
IMDb Votes
1.0
0.8
0.6
0.4
0.2
0.0
Arka Mediaworks
LBI Entertainment
Odd Lot Entertainment
Film 44
CBS Films
Excel Entertainment
StudioCanal
Working Title Films
Plan B Entertainment
Anonymous Content
Production House



Conclusion

Combining these insights, the ideal partners for new content would be those who have a proven record of both critical and popular success as seen in the graphs

Summary

1. Netflix's library has grown significantly over the years, with notable increase in number of movies & TV shows.
2. The analysis revealed that Netflix's catalog consists predominantly of movies compared to TV shows.
3. The data indicated that a large portion of the content available on Netflix was released in the last two decades.
4. When examining the availability of Netflix content across different countries, it was apparent that the United States had the largest selection of titles followed by Canada.
5. Through our analysis, we recognized that Netflix features a mix of films from both highly acclaimed directors and up-and-coming filmmakers.

DASHBOARD

NETFLIX

Data Insights on NETFLIX

CS418 PROJECT

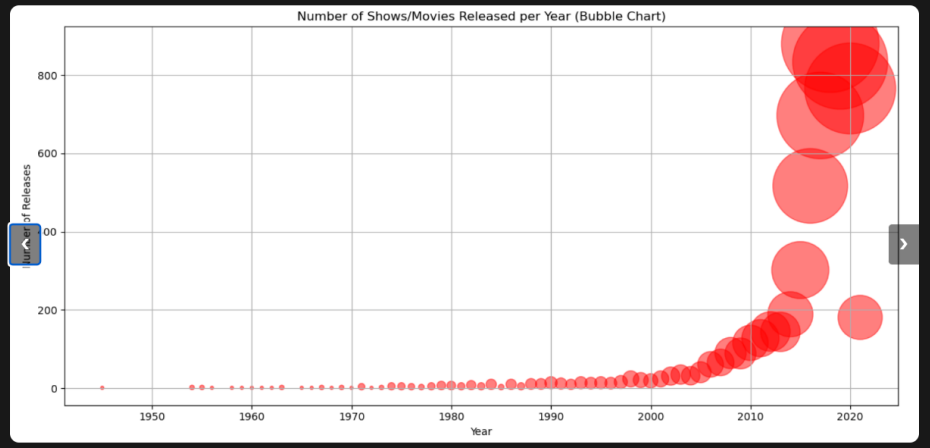
This project aims to uncover patterns in Netflix's vast library of films and TV shows, exploring how content types have evolved over time and identifying popular genres across different regions. Our analysis delves into viewer ratings, content distribution, and genre preferences to provide a comprehensive view.

Data Insights | Interactive Dashboard | IMDb Predictor

1 2 3 4 5 6

Number of Titles Added to Netflix

The bubble chart, line graph, and area graph all highlight Netflix's growth in adding shows and movies over time. These charts show that after the 1990s, there was a big jump in the number of titles Netflix offered, matching up with the popularity of streaming online. The bigger bubbles, rising line, and wider shaded areas in the graphs all point to how Netflix has been working hard to give viewers more choices, especially in recent years. This increase is likely because Netflix wants to meet market demand and technological advancements.



NETFLIX

Data Insights on NETFLIX

CS418 PROJECT

This project aims to uncover patterns in Netflix's vast library of films and TV shows, exploring how content types have evolved over time and identifying popular genres across different regions. Our analysis delves into viewer ratings, content distribution, and genre preferences to provide a comprehensive view.

Data Insights | Interactive Dashboard | IMDb Predictor

IMDb Predictor

Enter details to predict the IMDb score.

Movie: Rajamouli

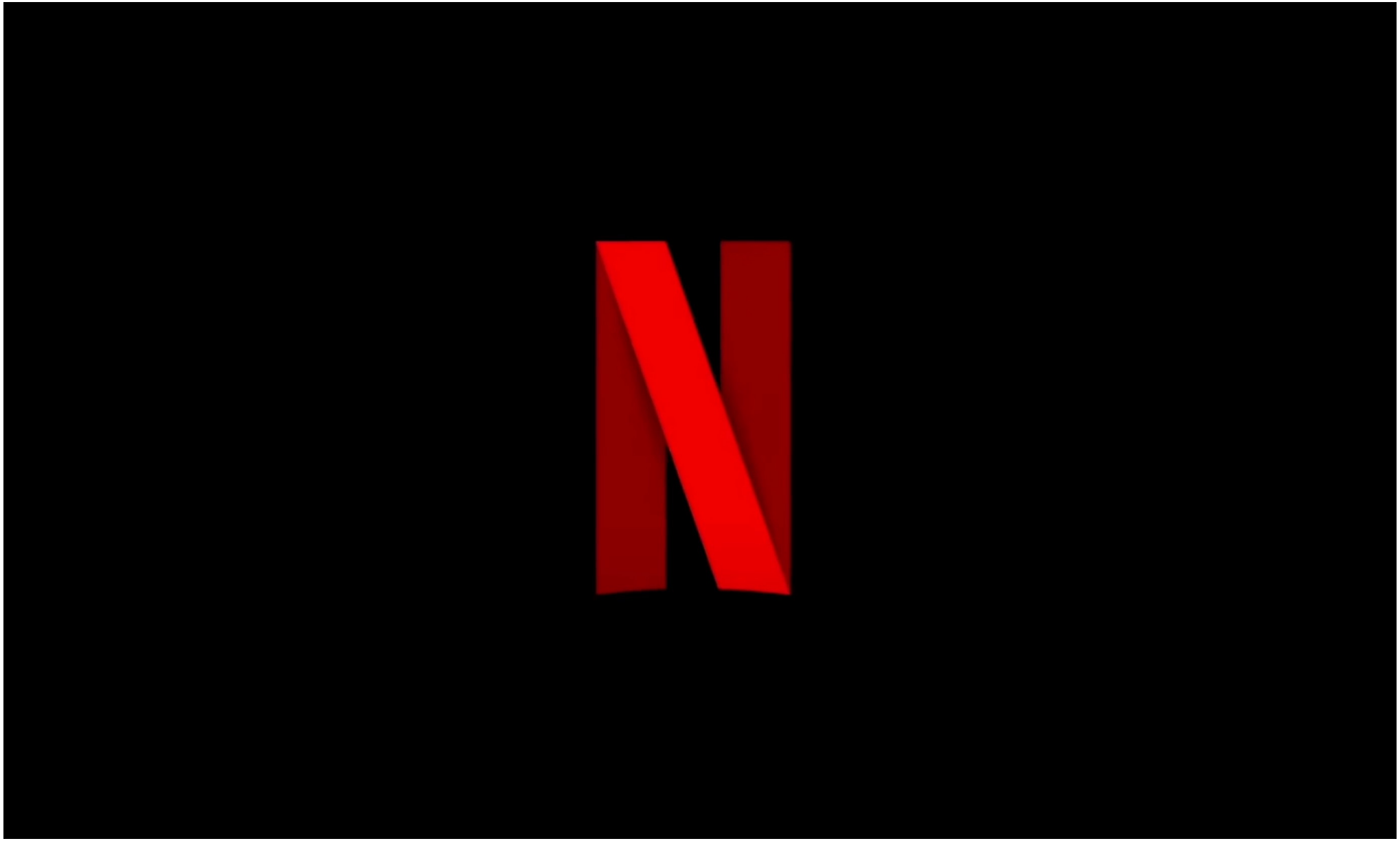
India

Action

Check

Predicted IMDb Score : 6.0090

DEMO



QUESTIONS ?

THANK YOU