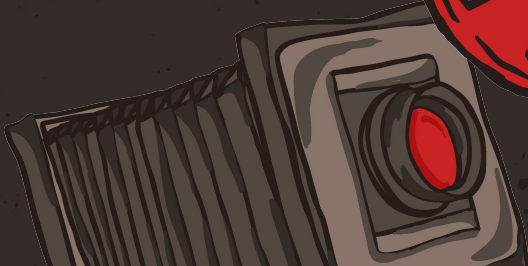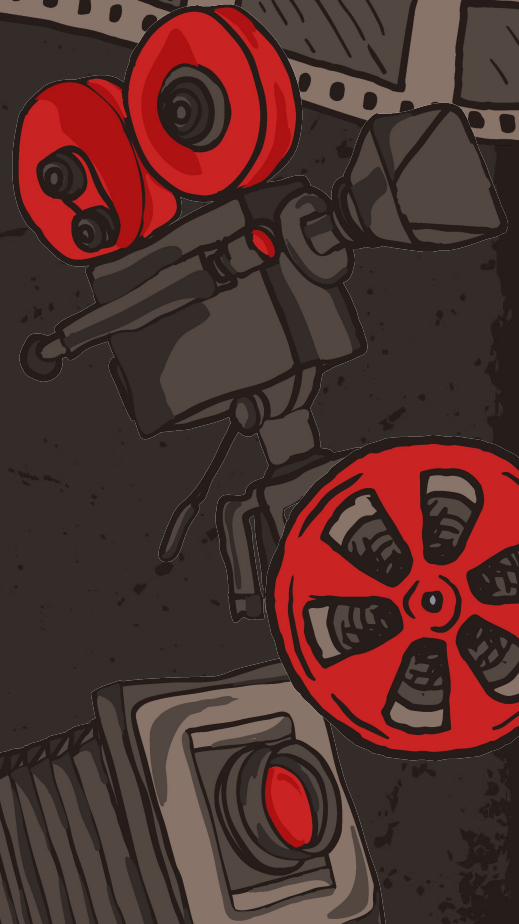STAT 405 - Data Science Computing Project

Factors Influencing Movie Ratings

Carter Jensen
Jackson Wegner
Michael Manley
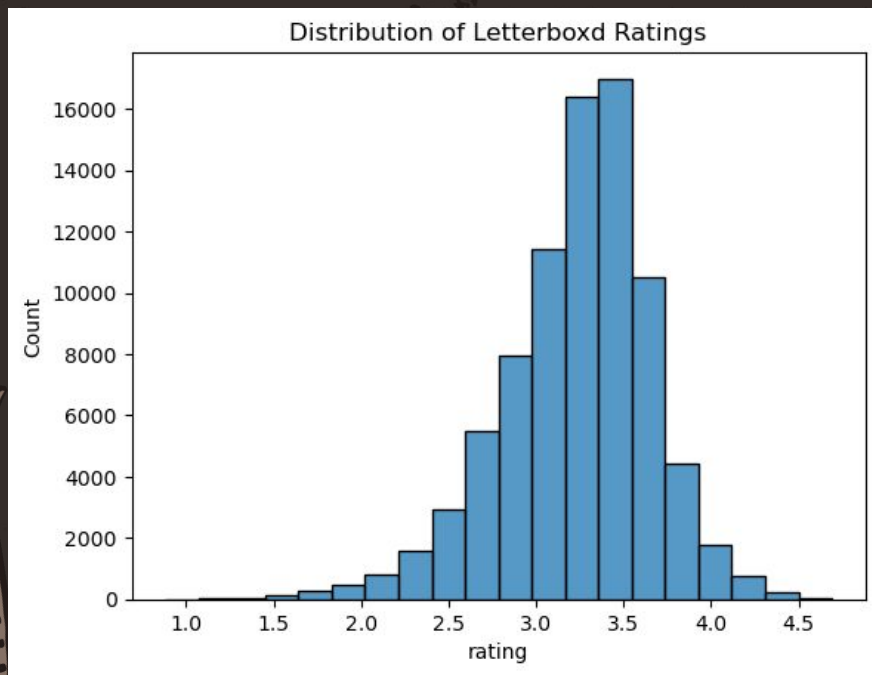Govind Kishan
Luis Granados

Overview

This project investigates the factors influencing movie ratings using a large Letterboxd dataset, which includes information on nearly 800,000 movies and with information on genre, rating, release date, actors, studio, and more..

Research Question

What factors have the greatest impact on a movie's rating?

- The film industry is influenced by multiple factors, but we wanted to identify what really drives a movie's rating. Does the genre, country of origin, or actor ratings have the largest impact?

Distribution of Letterboxd Ratings

The ratings are concentrated around the 3.0–3.5 range and the distribution appears normal. This suggests that the assumption of normality, which is necessary for our statistical modeling, is satisfied.

Parallel Computing on CHTC

We created
actors_and_ratings.csv
- **Would take about 8 hours run locally**
- **Required looping through 575602 actors**
- **Took about 16 minutes when run as 128 jobs in parallel**

Actor	Average Rating	Top rated film
Tom Cruise	3.37	Magnolia
Nicole Kidman	3.22	Big Little Lies
Timothee Chalomet	3.29	Dune: Part Two
Margot Robbie	3.14	About Time
Leo Di Caprio	3.49	The Departed

Statistical Methods Used

ANOVA

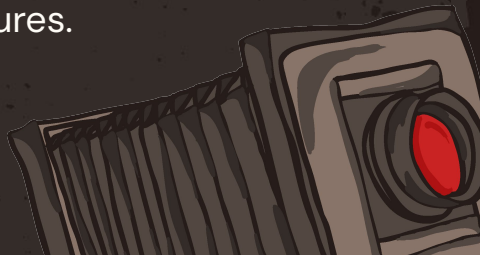– Analyzed how genre and country affect ratings.

Mean Aggregation

– Grouped by country and genre to identify the highest-rated groups.

Regression Analysis

– Explored how movie runtime influences ratings.

Machine Learning

– Used Gradient Boosting to predict ratings based on key features.

ANOVA Tests

Determine if ratings vary significantly by genre and country.

Genre:
- F-value = 1932
- p-value < 2e-16

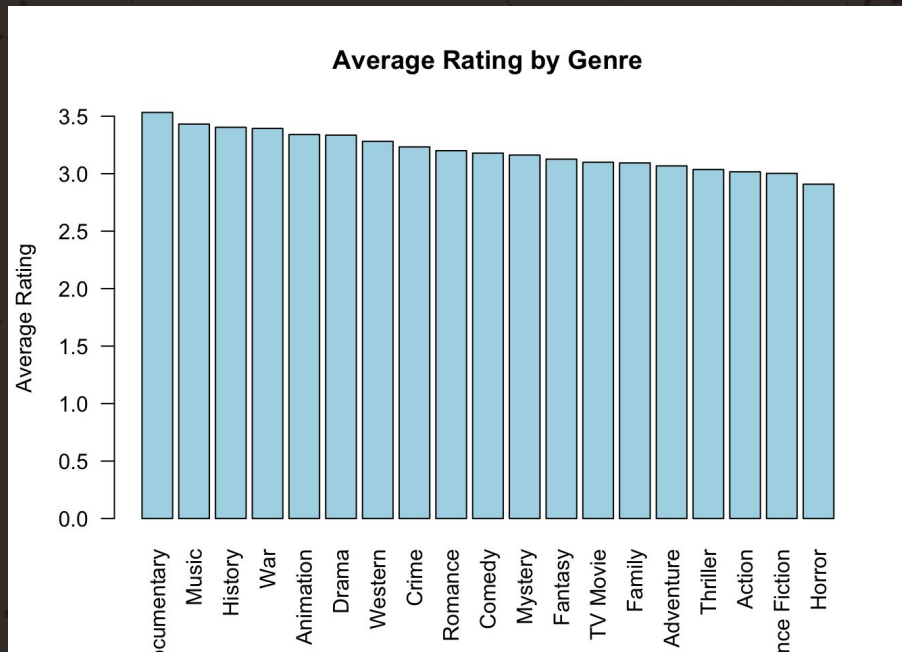**Country:**
- F-value = 87.18
- p-value < 2e-16

Both genre and country have a significant effect on movie ratings, highlighting the importance of these factors in determining how movies are rated by audiences.

Mean Aggregation - Genres

Top Genres

- Documentary (3.53)

- Music (3.43)

- History (3.40).

Mean Aggregation - Country

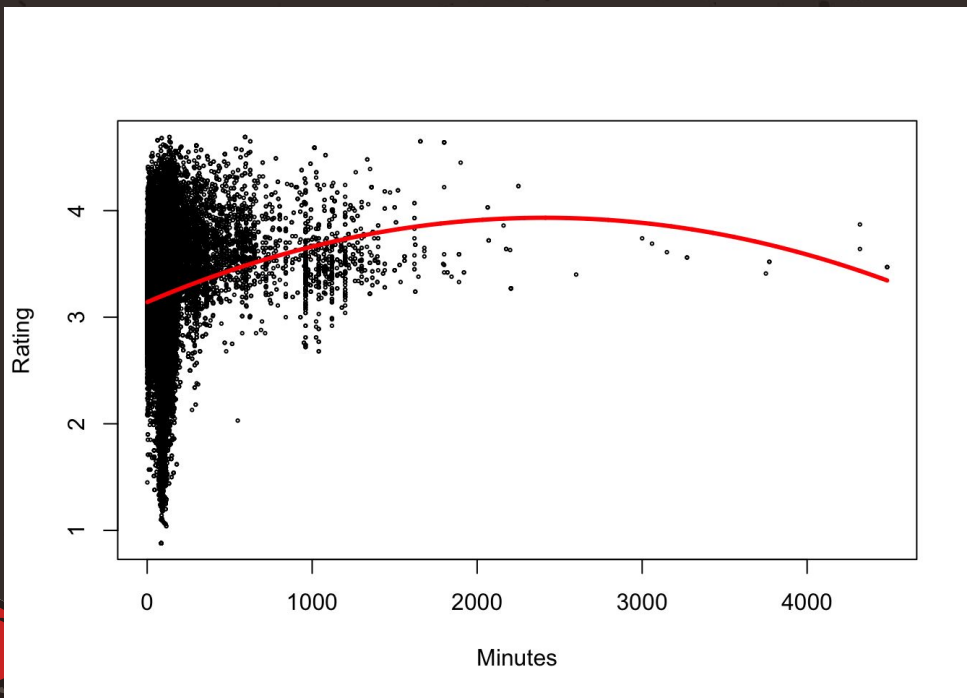Top Countries

- Libya (3.87)

- Martinique (3.79)

- Lesotho (3.78)

```
##                country   rating
## 107             Libya  3.866364
## 117        Martinique  3.790000
## 105           Lesotho  3.775000
## 46              Congo  3.720000
## 191            Uganda  3.707222
## 24            Bermuda  3.693333
```

Regression Analysis - Movie Length vs. Rating

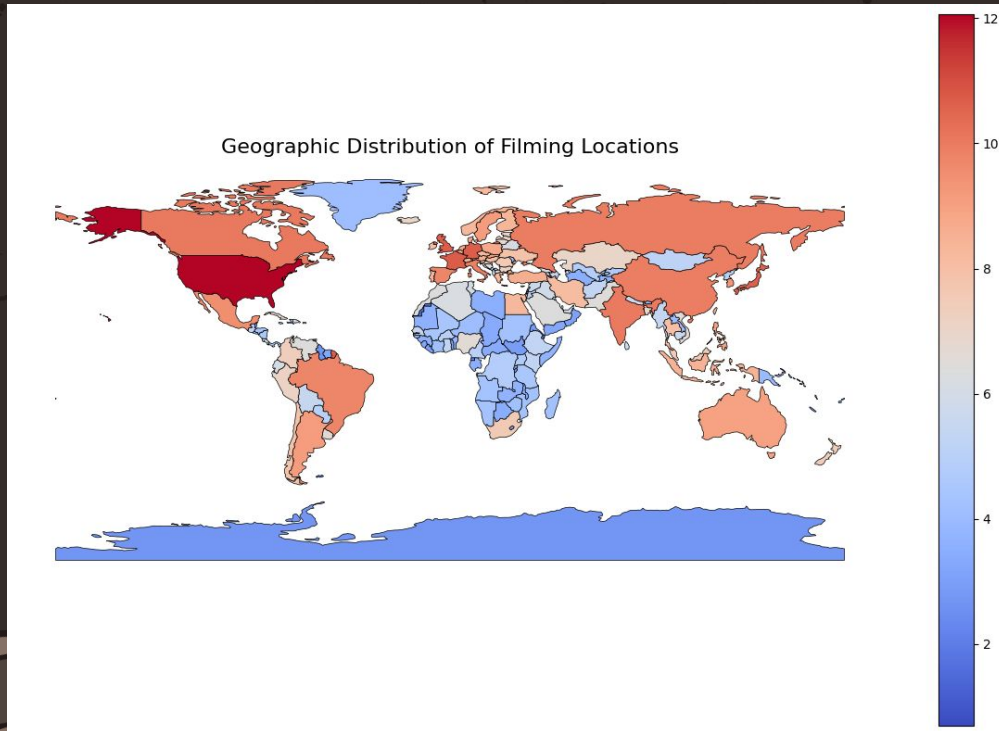Explore the relationship between movie runtime and ratings.

Quadratic regression revealed a non–linear relationship. Movies between **100–200 minutes had the highest ratings**, while both shorter and longer movies had lower ratings.

This suggests that there is an optimal movie length that tends to get higher ratings.

Geographic Distribution of Movie Locations

Investigate where movies are filmed around the world and identify trends in production locations
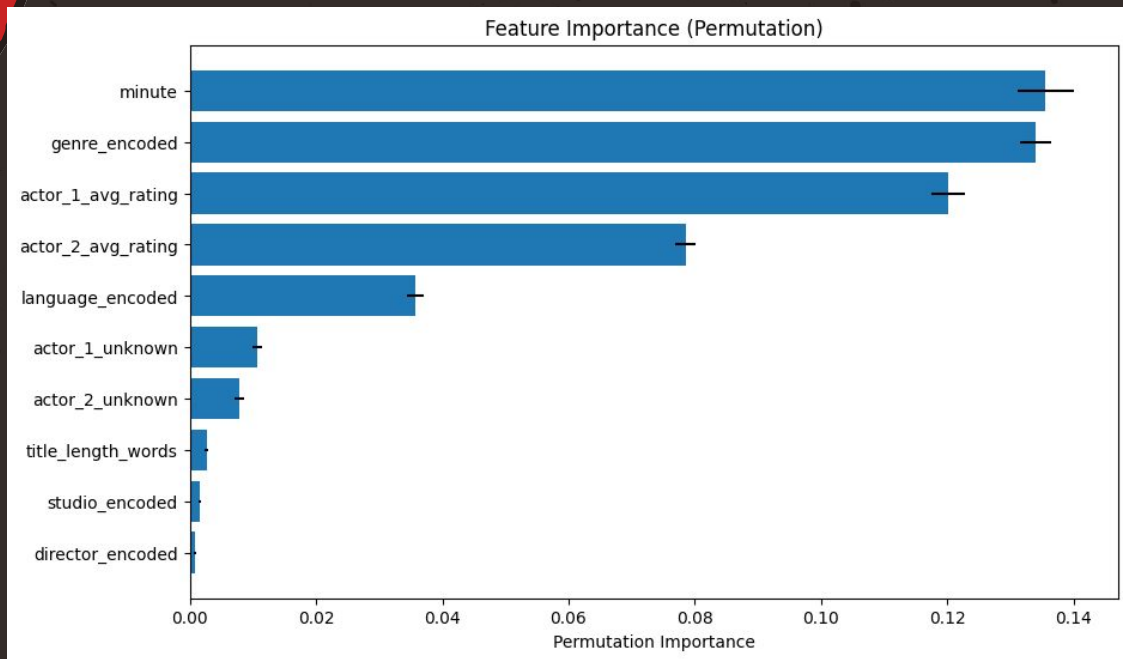

Geographic Distribution of Filming Locations

country	count
USA	174489
France	45725
UK	42914
Japan	41362
Germany	41325

Our results showed **The United States** has the highest number of filming locations, followed by **France** and **Germany**. These countries account for the majority of global film production.

Feature Importance - Predicting Ratings

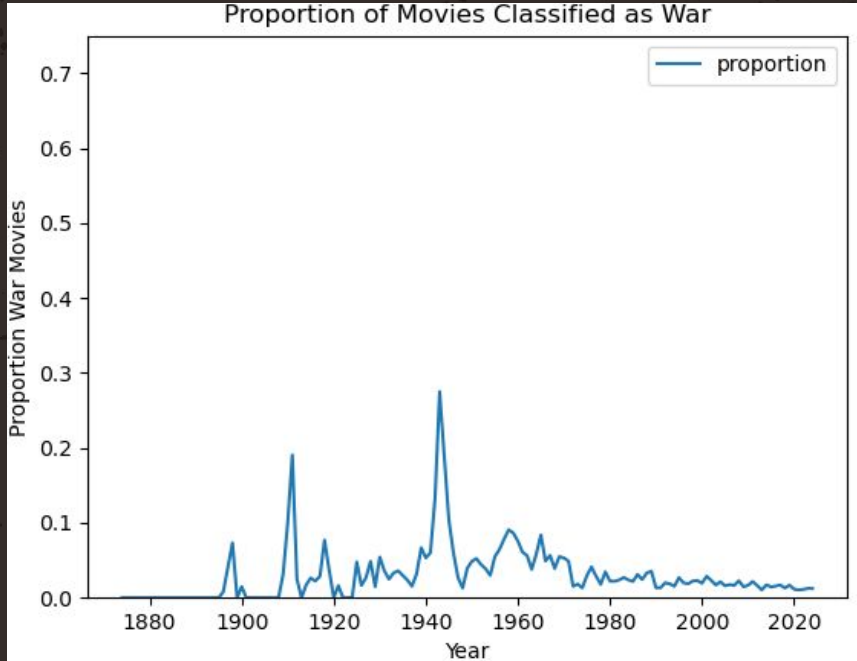Identify the most important features in predicting movie ratings.

Feature Importance (Permutation)

The most important features affecting ratings were **movie length**, **genre**, and **lead actor ratings**.

The Gradient Boosting Model was able to predict movie ratings with an R² score of 0.382, explaining 38.2% of the variance.
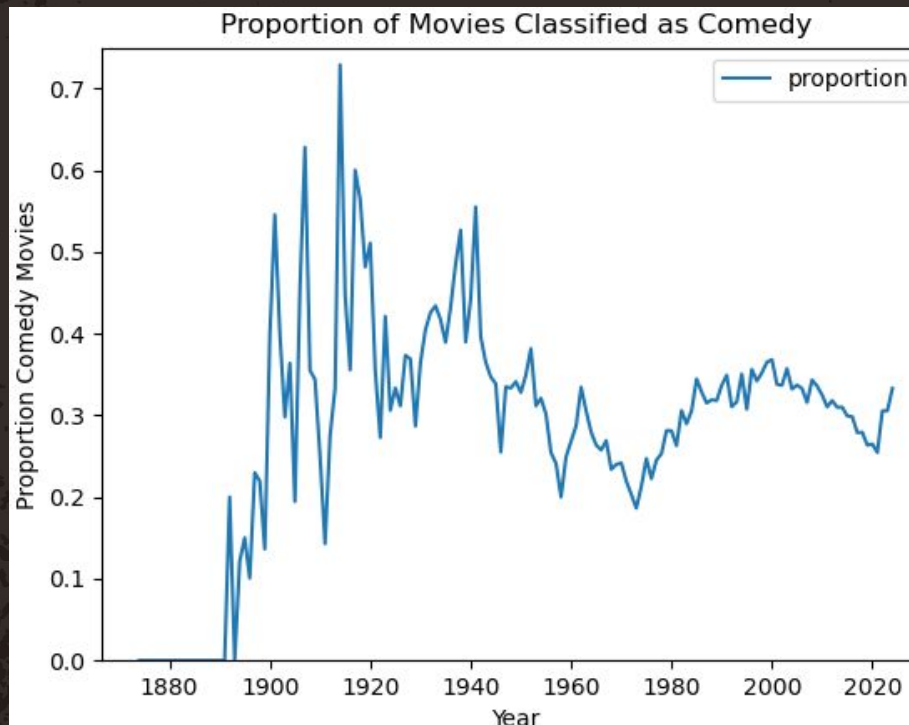
Genre Proportions Over Time

Additionally, we analyzed how genres have changed over time
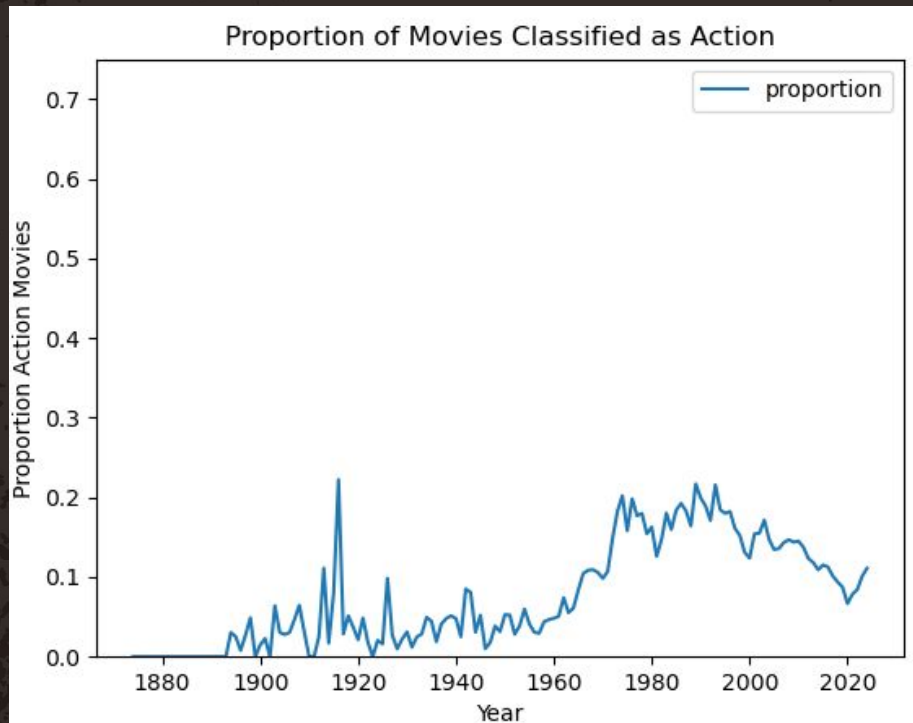

The plots highlight shifting audience preferences and broader industry trends across genres in their respective time periods.

War films spiked during WWII and the post–war era, reflecting wartime experiences, with another surge during the Vietnam War in the 1970s.

Proportion of Movies Classified as Comedy

Comedy films peaked in the 1910s–1920s with silent film stars like Charlie Chaplin, then surged again in the 1970s with the rise of slapstick and family-friendly humor.

Proportion of Movies Classified as Action

Action films saw a sharp increase from the mid–1980s, driven by blockbuster franchises like Die Hard, Terminator, and superhero films, which became a major global draw.
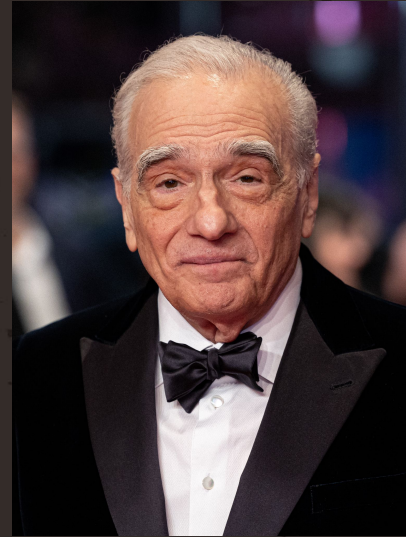
Conclusion & Next Steps

Key Findings:
Genre, runtime, and lead actor ratings are the primary drivers of movie ratings, with geography also influencing film success across regions.

Next Steps:
Future research could explore the impact of marketing, director reputation, and budget on movie ratings and success.

Thanks!