CinemaScore Predictor: An AI Model for IMDb Score Prediction

Fredy Aguirre
Ketan Parekh
Brian Hansen-Turton
Alan Hunt

Concept and Motivation

Challenges in the Film Industry

- Success in the competitive film industry requires a strategic understanding of audience preferences and movie dynamics.

Development of CinemaScore Predictor

- An innovative tool designed to forecast IMDb scores and provide actionable insights into audience reception.

- Helps identify key elements for predicting a movie's critical and financial potential before production begins.

Strategic Team Investment

- Leveraged the team's expertise and financial success from a previous project to develop this groundbreaking solution.

Value of IMDb Scores

•High IMDb scores reflect strong audience engagement and satisfaction.

Key Predictive Factors

•AI leverages factors such as director, actors, genre, and rating to forecast IMDb scores.

Actionable Insights

•The CinemaScore Predictor provides filmmakers with data-driven insights for crafting audience-focused films.

Empowering Filmmakers

•Equips producers and filmmakers with advanced tools to make informed decisions.
•Aims to deliver memorable cinematic experiences that resonate with audiences.

Data Exploration and Cleanup

Data Exploration

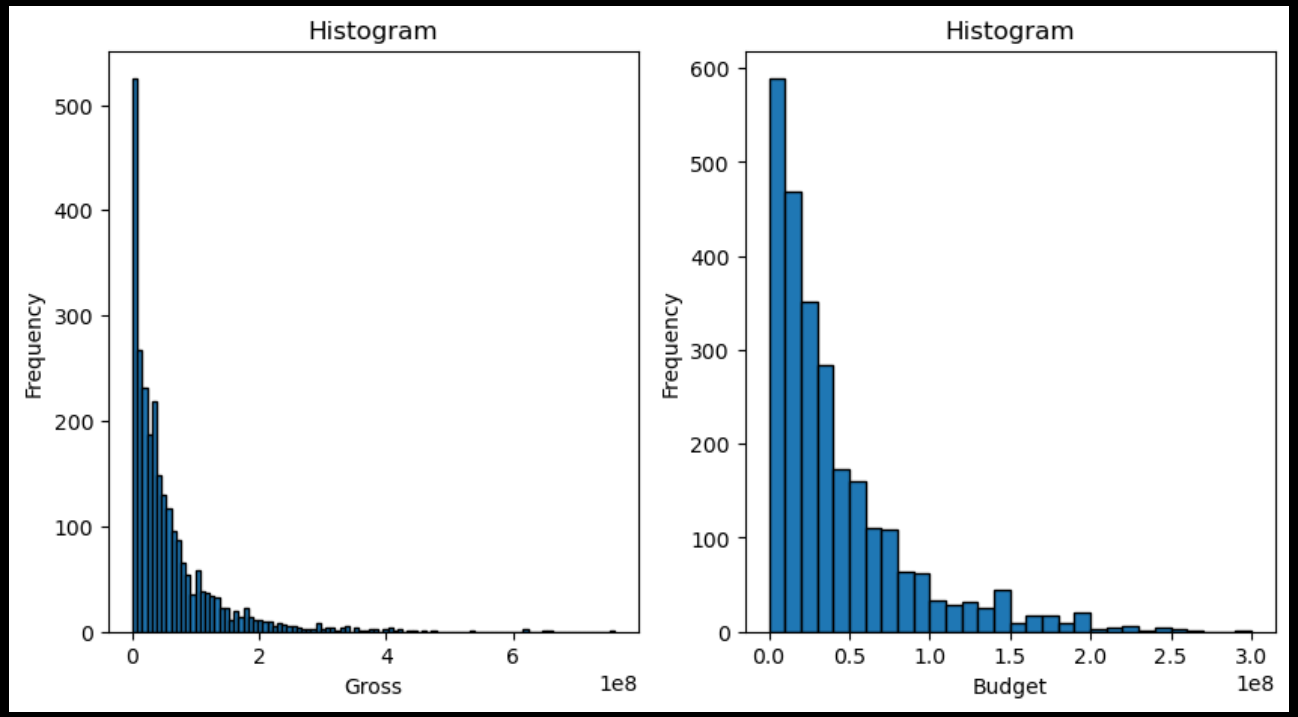
Exploratory data analysis (EDA) is performed to understand patterns and correlations within the dataset. Techniques such as visualization and summary statistics are used to gain insights into the data's structure and content.

Data Cleanup

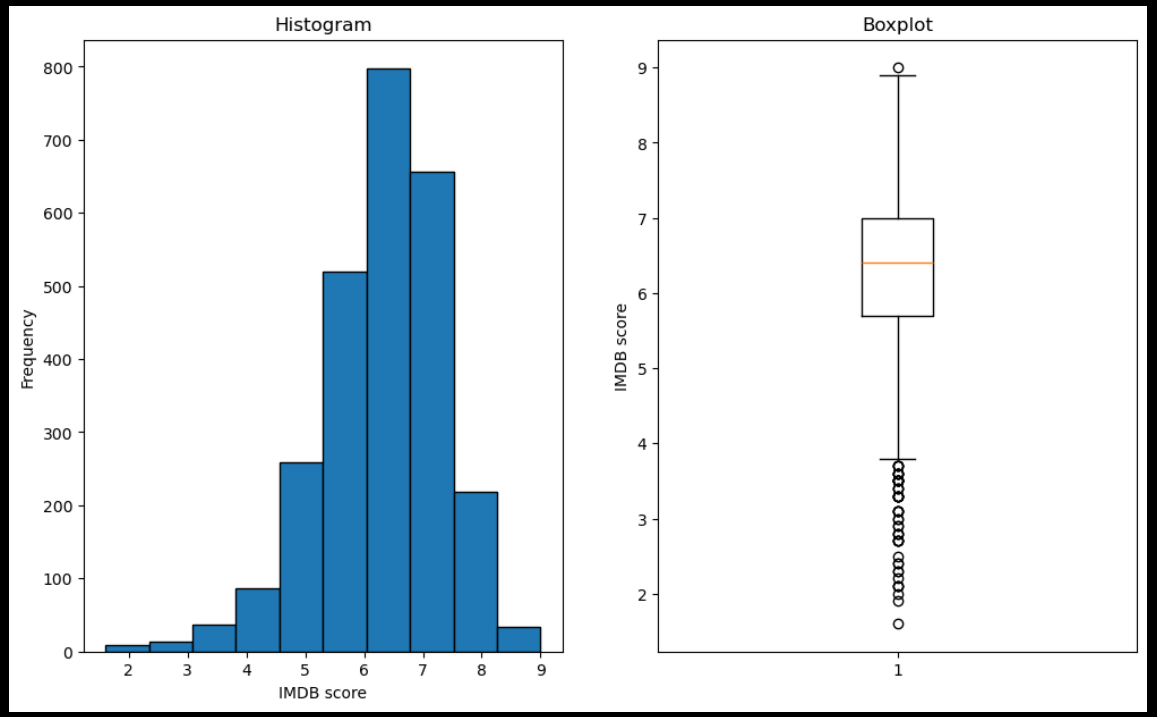
Data cleaning involves removing duplicates, handling missing values, and standardizing data formats. This step is crucial to ensure data accuracy and reliability for model training.

Cleanup Challenges

Missing Data

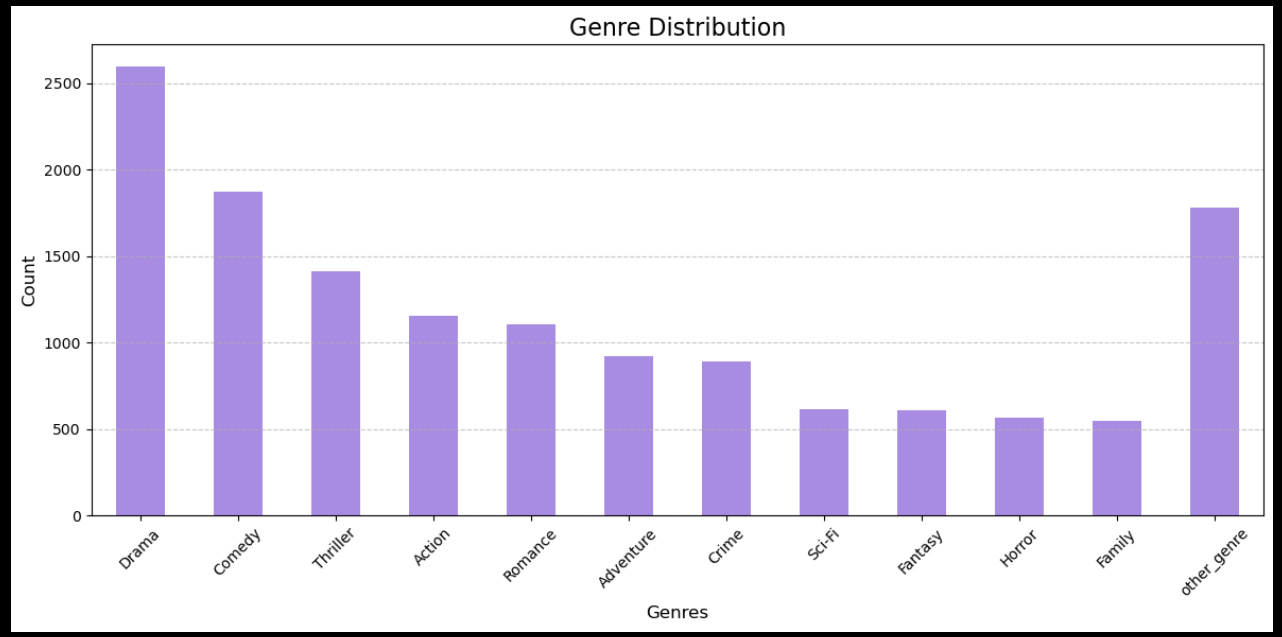


Outliers

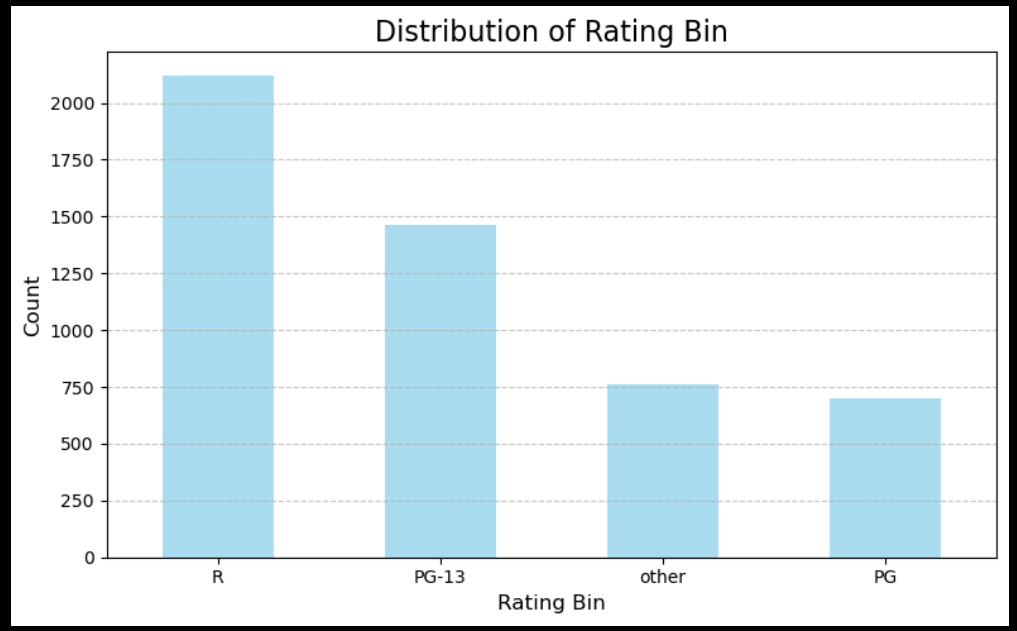


Cleanup challenges

- Genre



- Content Rating



Cleanup Challenges

Missing Values

- Actor_1_Name
- Actor_2_Name
- Actor_3_Name
- Director

Filter

- Gross Budget
- Language
- Country

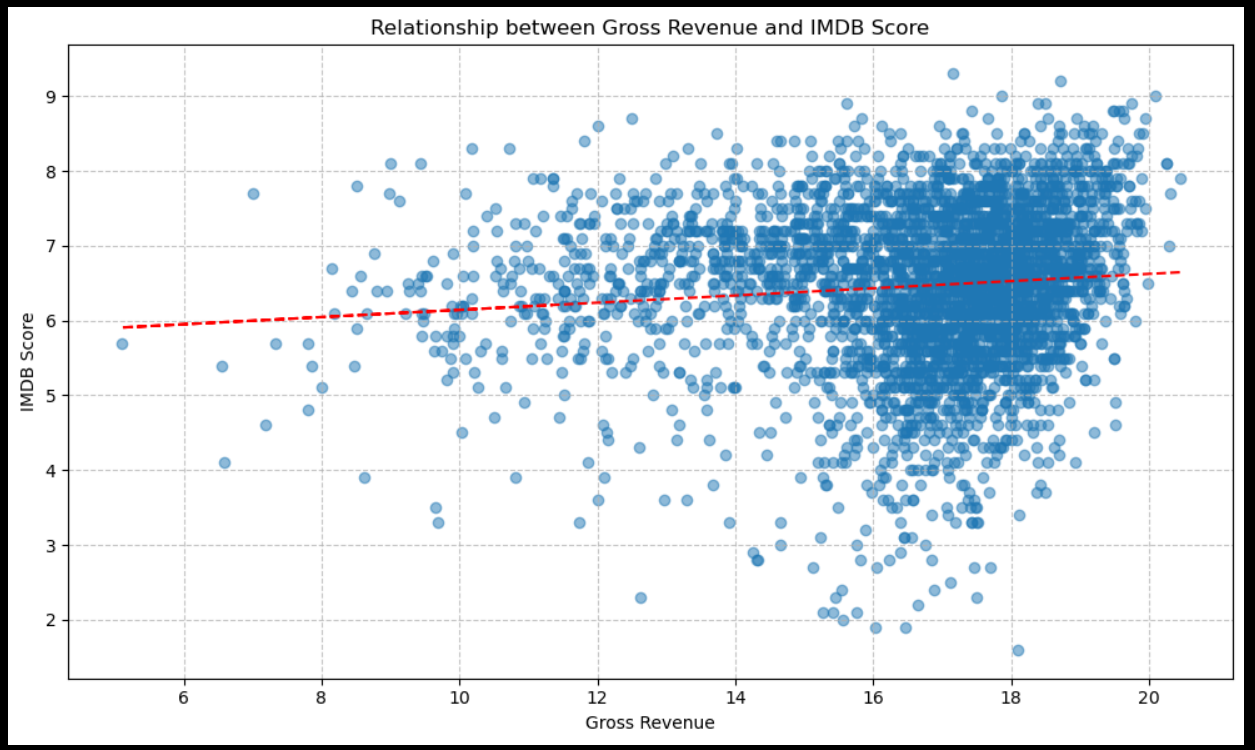
Drop Columns

- Movie Title
- Movie URL
- Plot keywords

Modeling

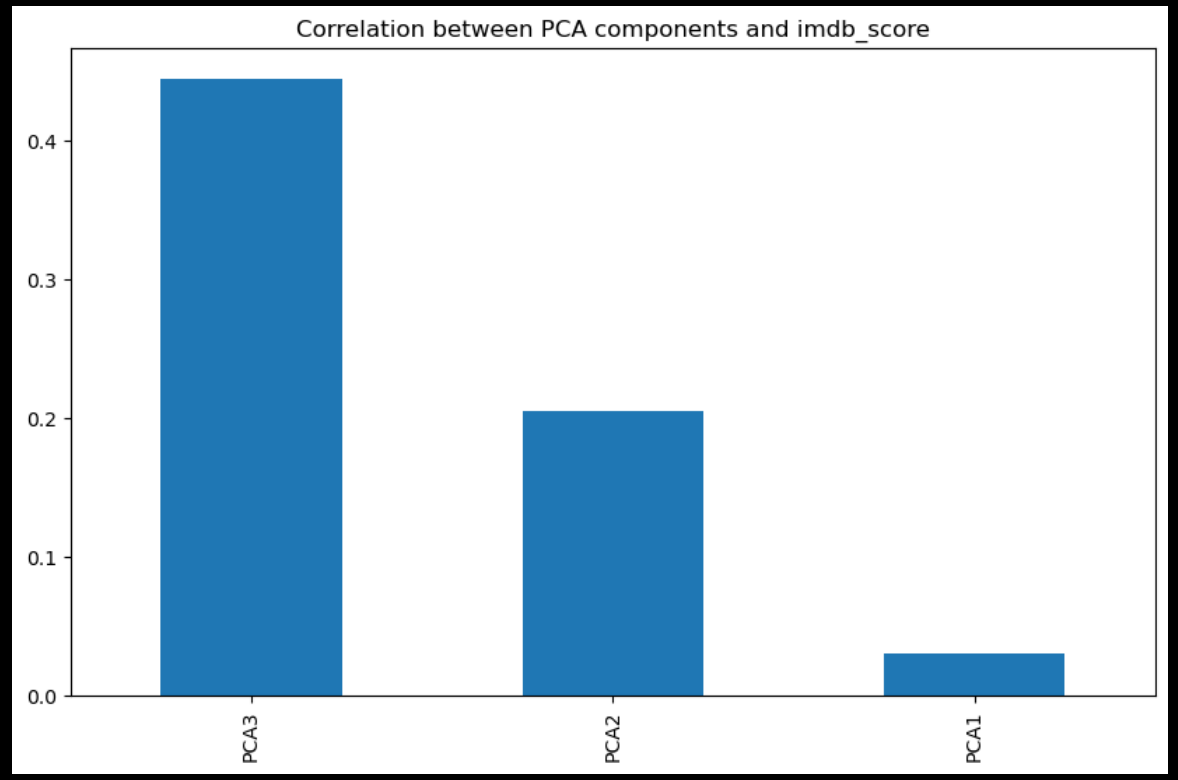
Initial Findings

Revenue Trend with Score

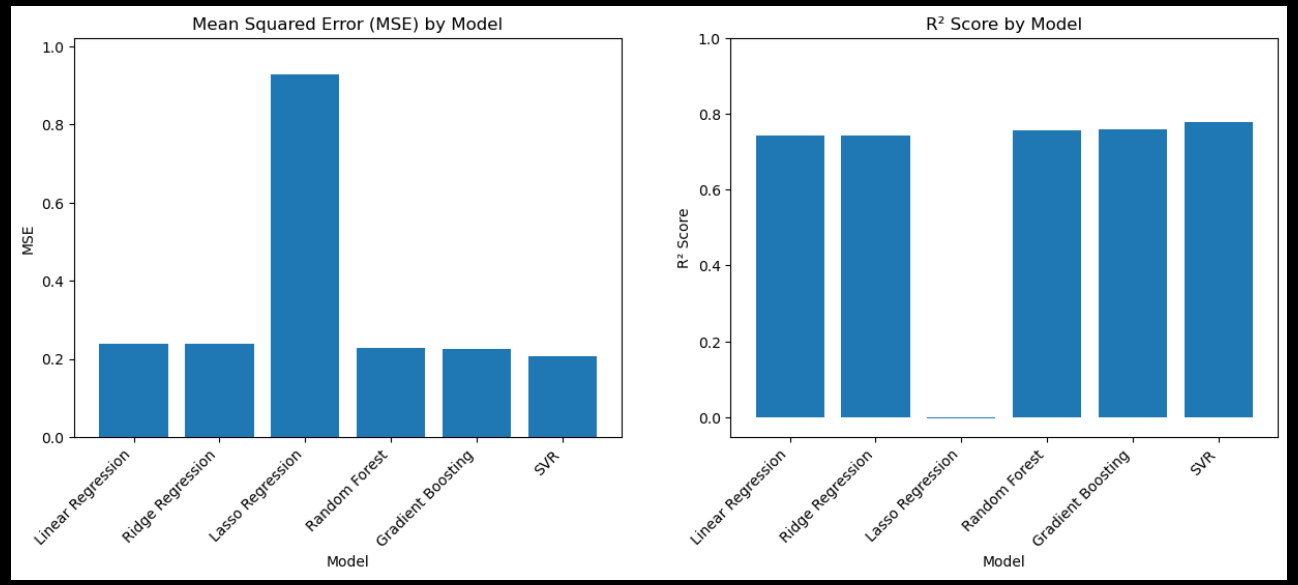


Modeling

Correlations and PCA



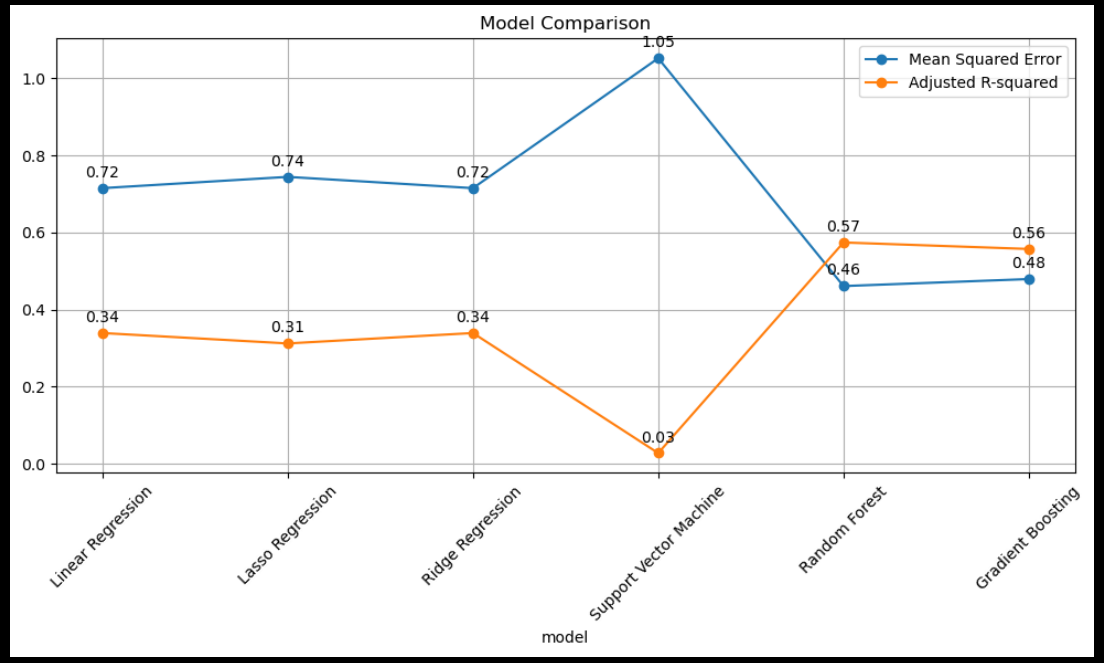
Model Performance



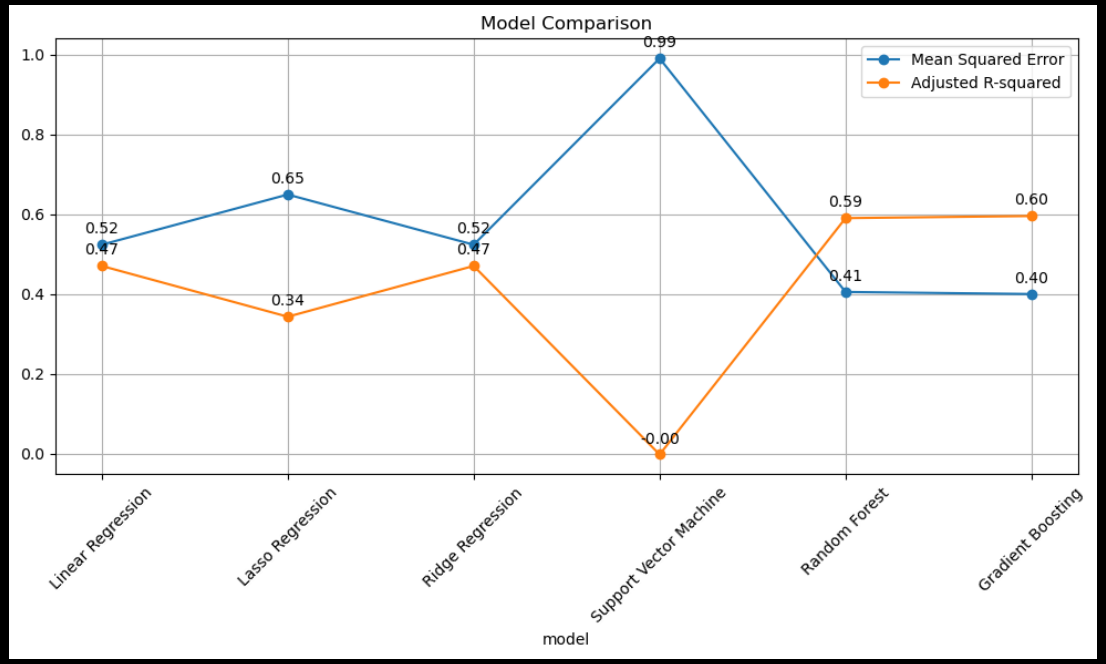
- Pre-data processing and PCA
- No apparent linearity
- Random Forest, Gradient Boost and SVR looked promising

Modeling

Initial Model Analysis



Post Data Processing



Modeling

Linear Regression Metrics

Mean Squared Error	0.523458
R Squared	0.497103
Adjusted R Squared	0.470216

Gradient Boost Regressor Metrics

Mean Squared Error	0.398399
R Squared	0.617250
Adjusted R Squared	0.596786

Next Steps: Future Research and Development

Research Areas

- **Use social media sentiment analysis for better predictions.**
- **Assess international market influence on scores.**

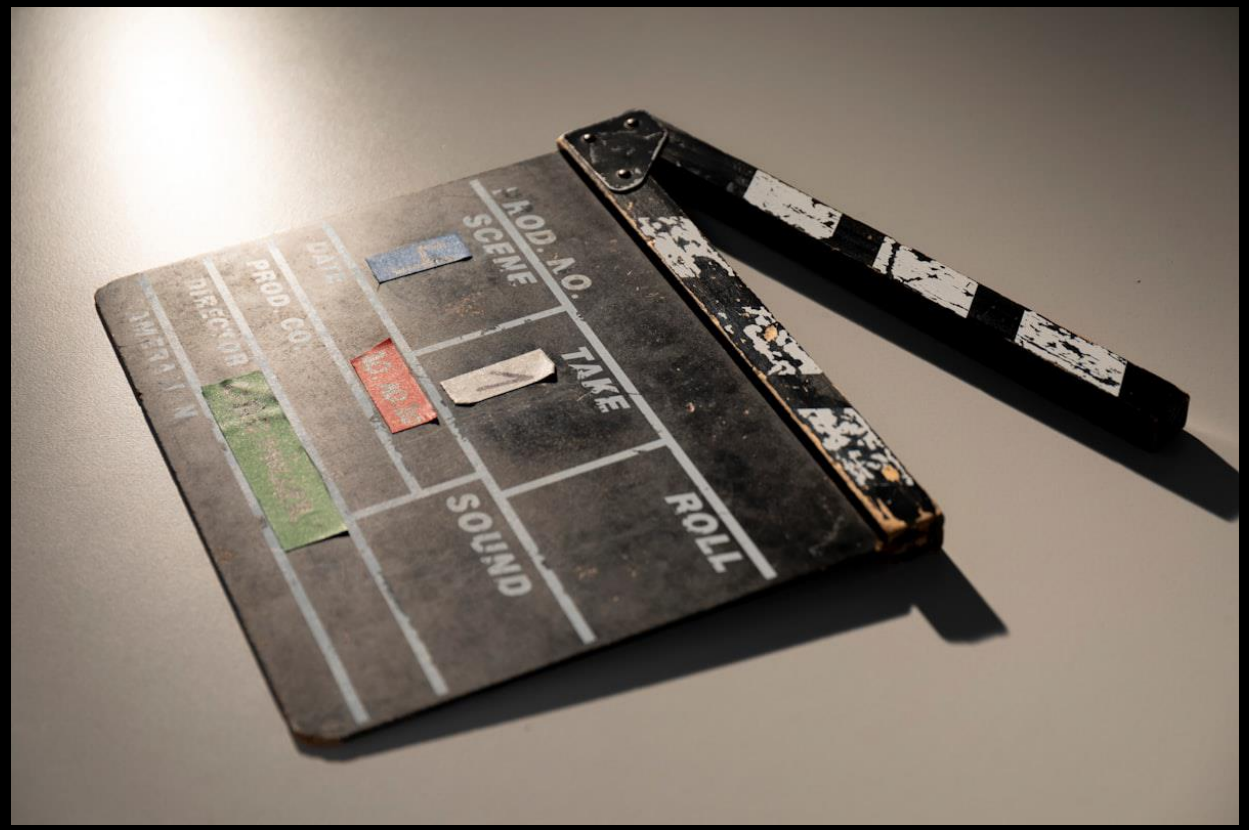
Development Plan

- Broaden film datasets with more IMDb data.
- Add a feedback loop for real-time model updates.
- Partner with experts to validate predictions.
- Focus on niche genres and global films.

GitHub Repository

[Explore the CinemaScore Predictor Code]

Venture Capital Inquires



Acknowledgments

Suzanna Ayash
For her unparalleled brilliance, destined to one day earn a Nobel Prize in Economics.

Jimmy Tran
For his unwavering dedication and exceptional guidance in leading us through this class.

Honorable Mention

John Ellis
For effortlessly winning the admiration and affection of our group.

CinemaScore Predictor Demonstration