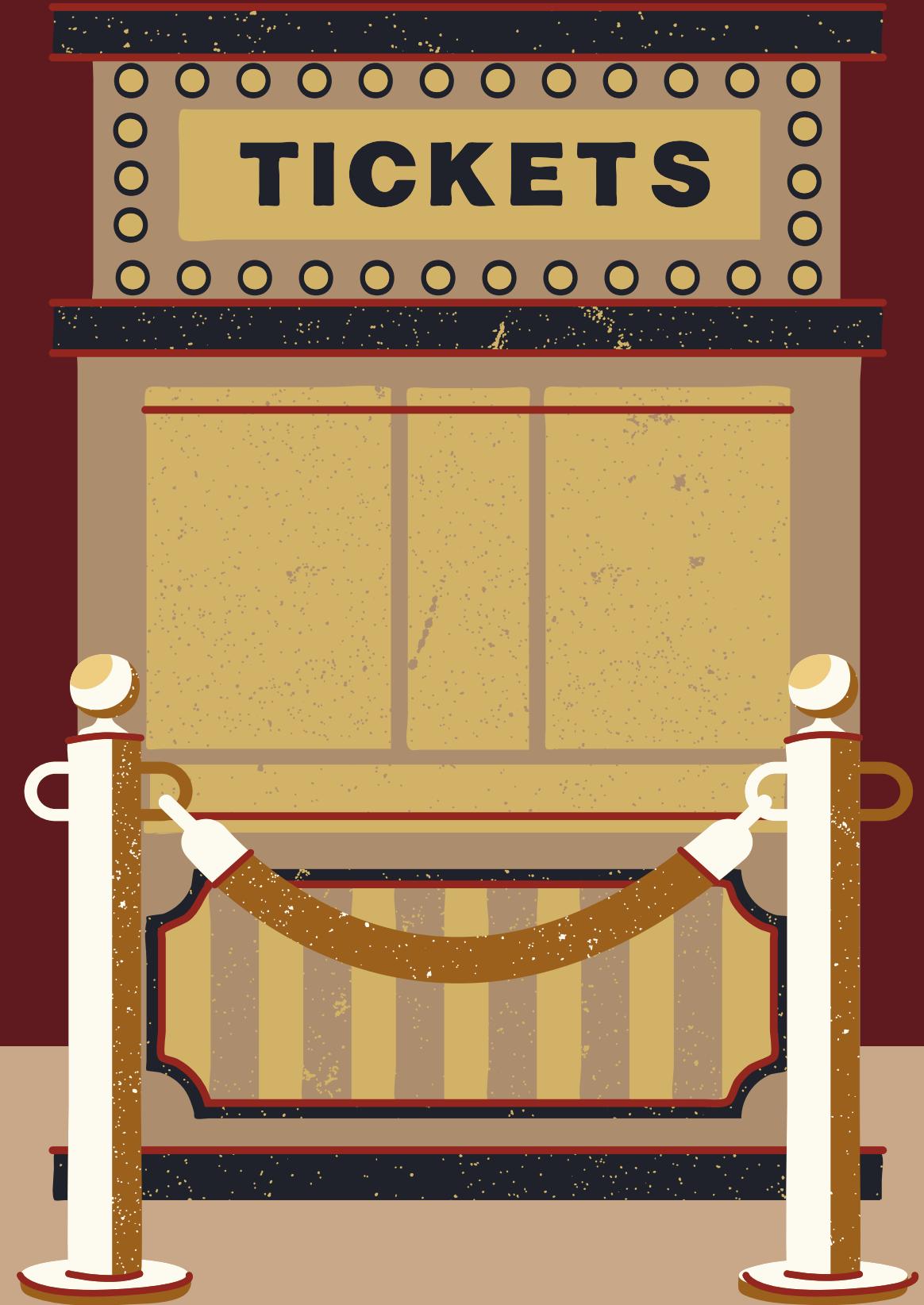




# THE NEXT BEST MOVIE



BROUGHT TO YOU BY : FCVCF



# Objective

Predict the next best-performing movie based on historical data

Create a predictive model that accurately estimates:

Potential revenue for upcoming movies.

Identifies key factors: budget, genres & release timing.

# Potential Impact

Movie studios can use these insights to optimize their investments.

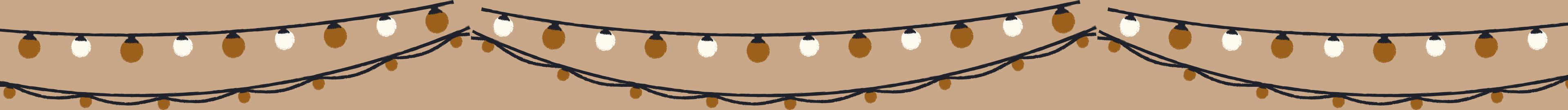
Align creative efforts with audience preferences, ensuring higher engagement.

Empower stakeholders to make informed decisions.



# DATA SELECTION

FCVCF Productions



# FULL TMDB MOVIES DATASET 2024



Over 1 million movies



Release dates



Revenue



Titles

# DATA CLEANING

- Reduced the irrelevant data by dropping columns
- Sorted movies by revenue to prioritize "best sellers"
- Kept only the last 20 years and filtered the outliers



# FEATURE ENGINEERING

- Scaled numeric variables using MinMaxScaler
- Split the release\_date column into release\_month and release\_day
- Normalized and split the genres column
- Extracted trends in movie titles (keywords)
- Tokenization: Breaking down text into individual words
- Feature Representation: Extract meaningful patterns from raw data



# MODEL BUILDING & EVALUATION

## Revenue Prediction:

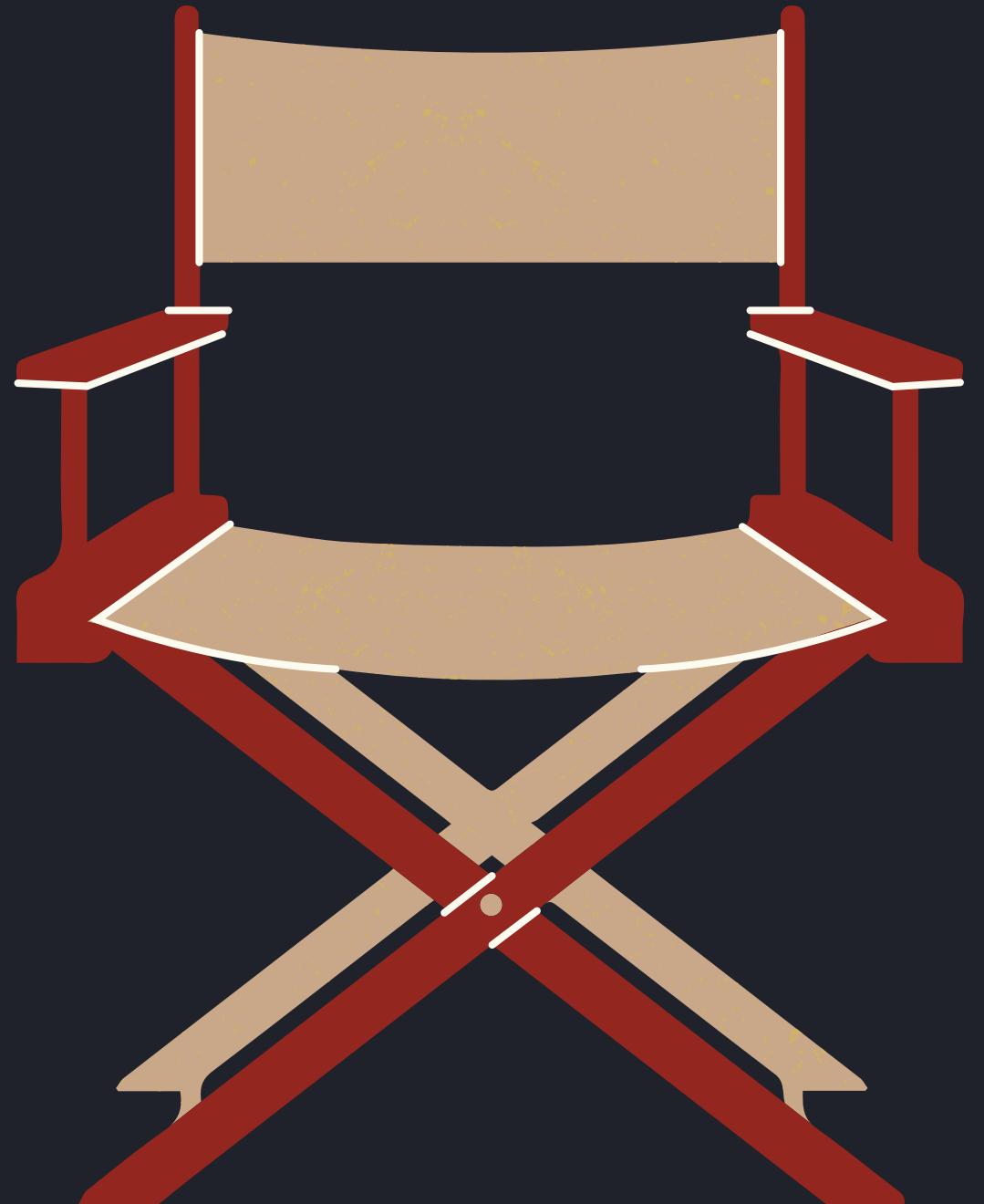
- RandomForestRegressor
- Input Features: budget, runtime, release\_year
- Evaluation Metric: Mean Absolute Error (MAE), achieving \$X (value)

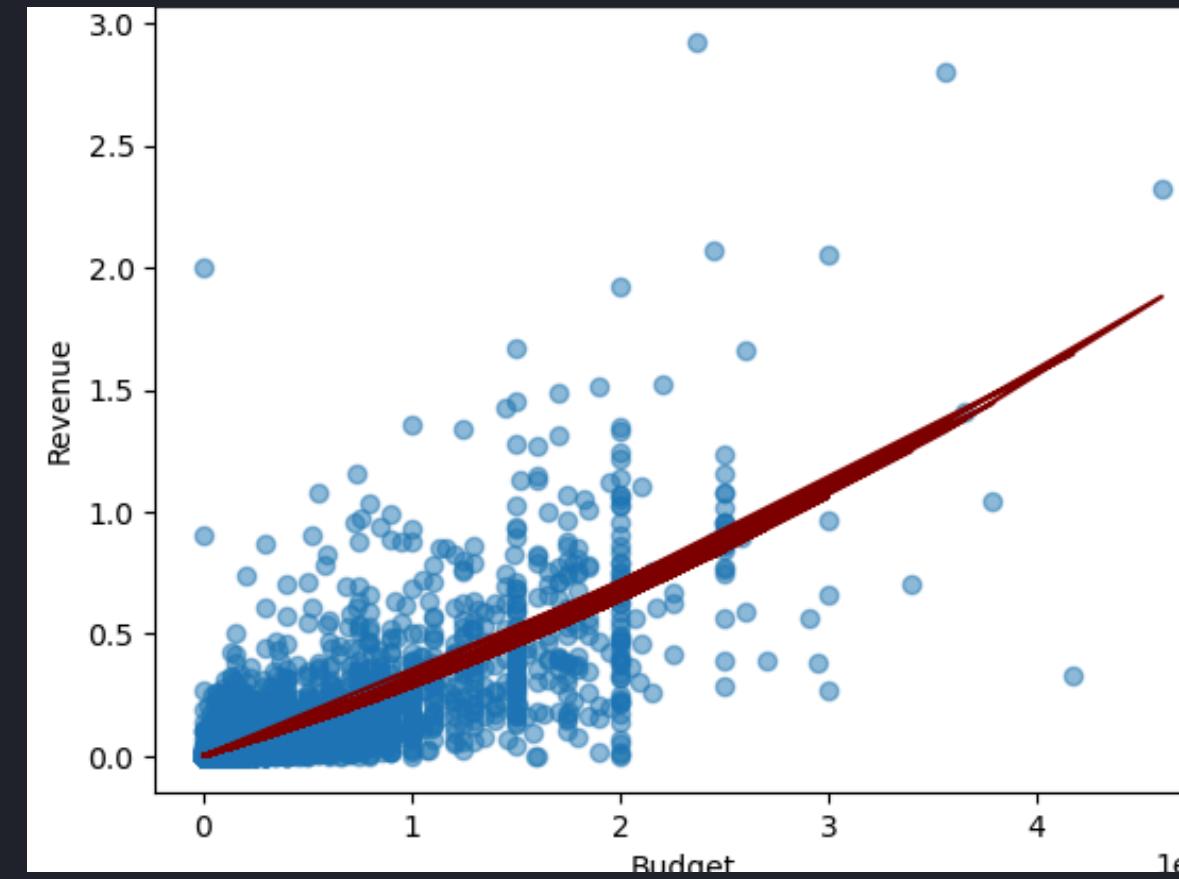
## Genre & Date Classification:

- RandomForestClassifier
- Input Features: release\_date, genres
- Evaluation Metric: Accuracy, achieving X value

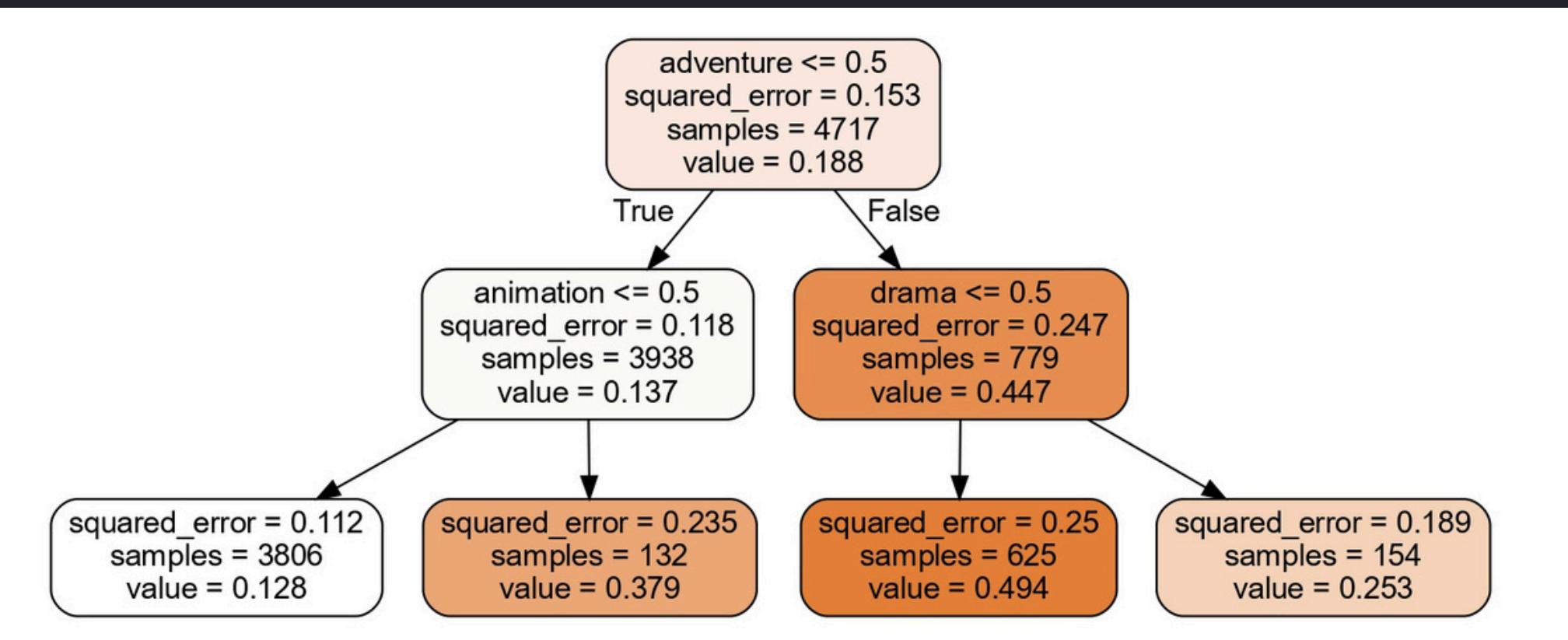
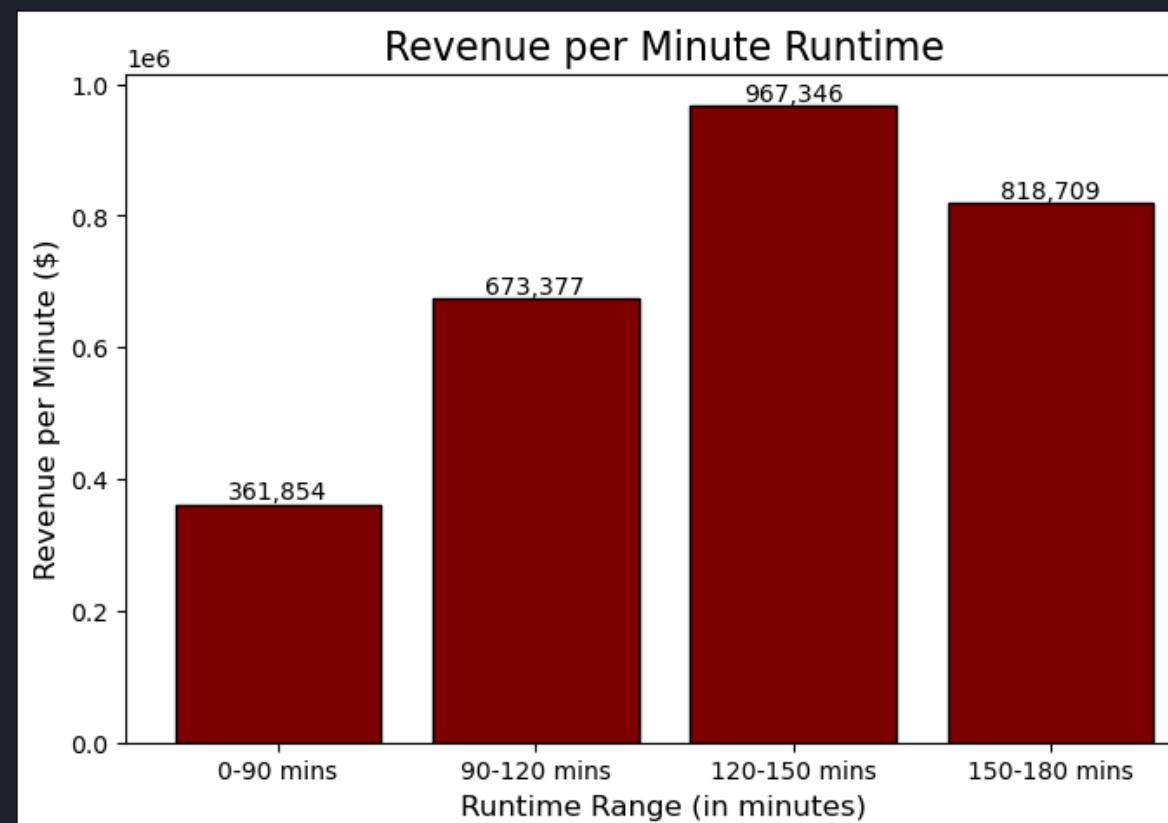
# VALIDATION

- Train-Test Split: Ensured an 80/20 split for training and testing datasets.
- Cross-Validation: Validated consistency in model performance across different folds.
- Random Forest Classifier
- Linear regression model





# HYPERPARAMETER TUNING & MODEL OPTIMIZATION



# KEY FINDINGS & INSIGHTS



**Revenue Drivers:** High budgets and longer runtimes correlate with higher revenue but show diminishing returns beyond a threshold, with smaller budgets and moderately long runtimes (120–150 minutes) often being more efficient in revenue generation.



**Genre Trends:** Adventure, science fiction and action family genres dominate top revenue earners.



**Title Impact:** Short, impactful titles with emotional or intriguing keywords perform better.



# REAL WORLD USAGE

Data-driven insights to optimize budgets, runtime, and marketing strategies.

Fimmakers

Predict genre combinations and titles likely to succeed in upcoming years.

Studios

Reduce financial risks by forecasting revenue potential.

Investors





# REAL WORLD USAGE

Revolutionizing the filmmaking process by combining creativity with analytics.

Process

Avoid reinforcing stereotypes or limiting diversity in storytelling.

Ethics



# CHALLENGES

- Understanding the data
- Tool familiarity
- Model selection and preparation
- Interpreting metrics
- Time and resource limitations
- Git s\*\*t



# FUTURE WORK & IMPROVEMENTS

- Test More Algorithms: Compare performance to find the best fit.
- Use Cross-Validation: Ensure model robustness across data splits.
- Automate Tuning: Optimize hyperparameters for better accuracy.
- Expand Dataset: Add demographics, competitor data, and social sentiment.

# BUDGING TITLES

- Georgia Have Huevos
- Have Sick Oz
- Spies Hacksawclub
- Lonely Atomic Luna
- Kazakhstan Serenity Everywhere
- Shaun Ultimate Picture
- Wakanda Straight Arrow
- Kitties Foreigner Cleaning
- Fantasy Margin Hidalgo
- Nebraska Inferno Aftershock



*A bootcamp to remember*

Don't hesitate to ask any questions!

