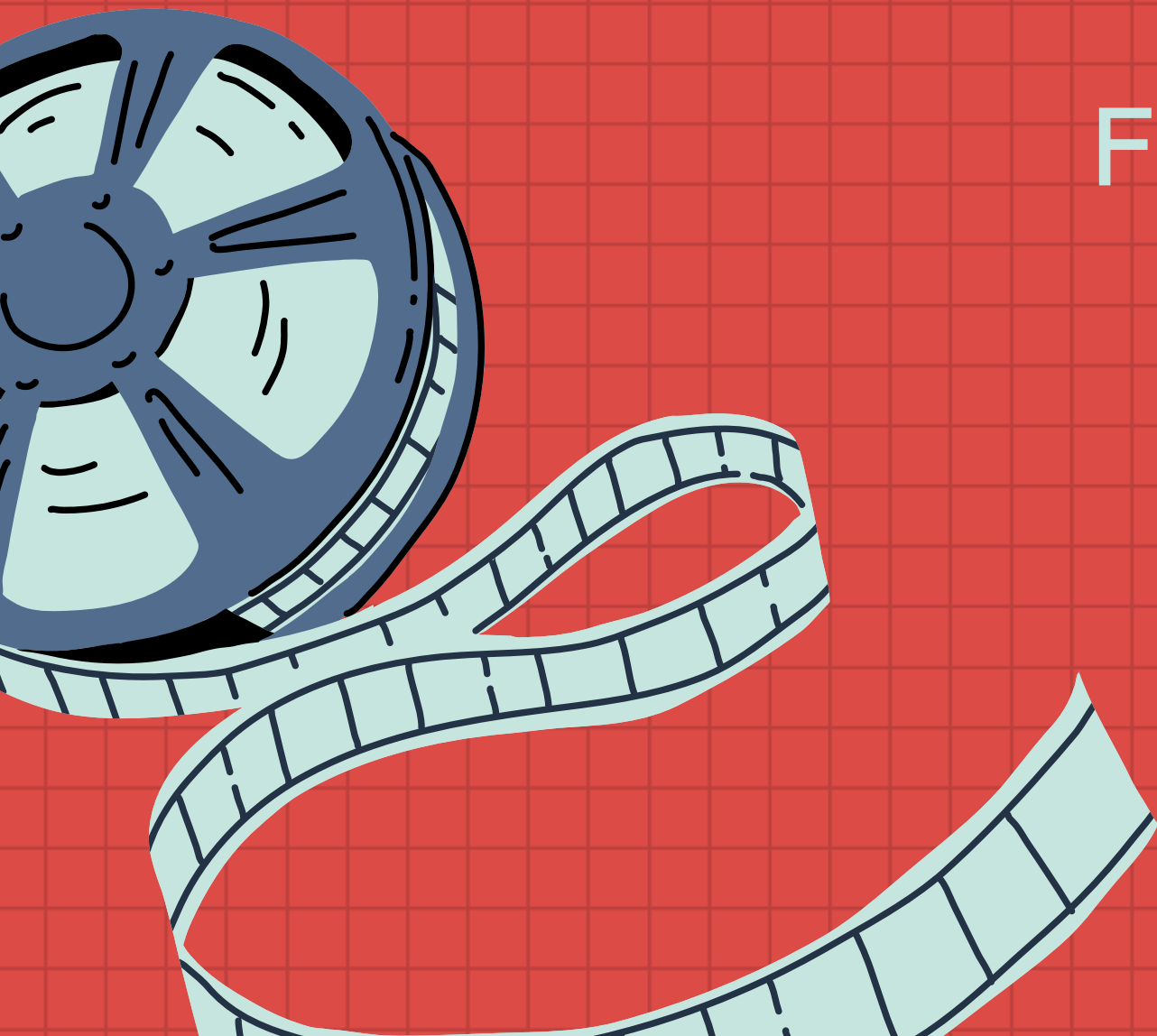
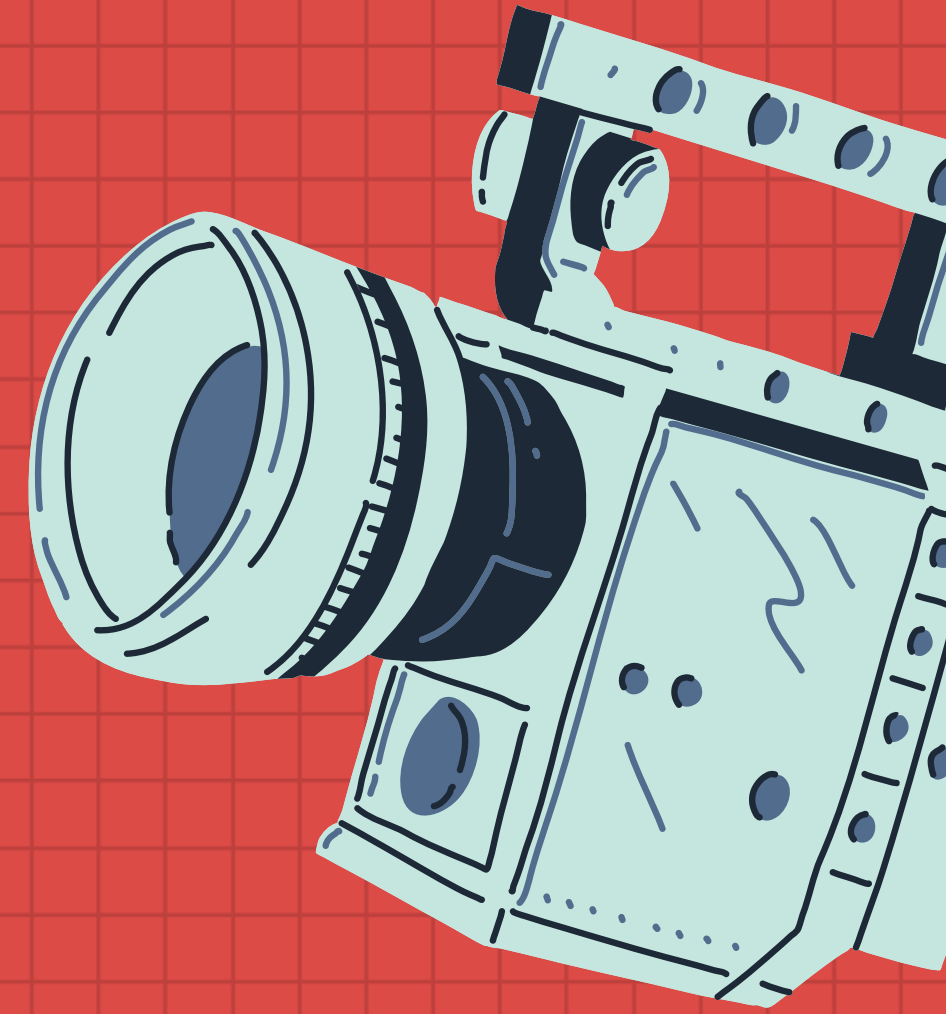




The Overthinkers present:

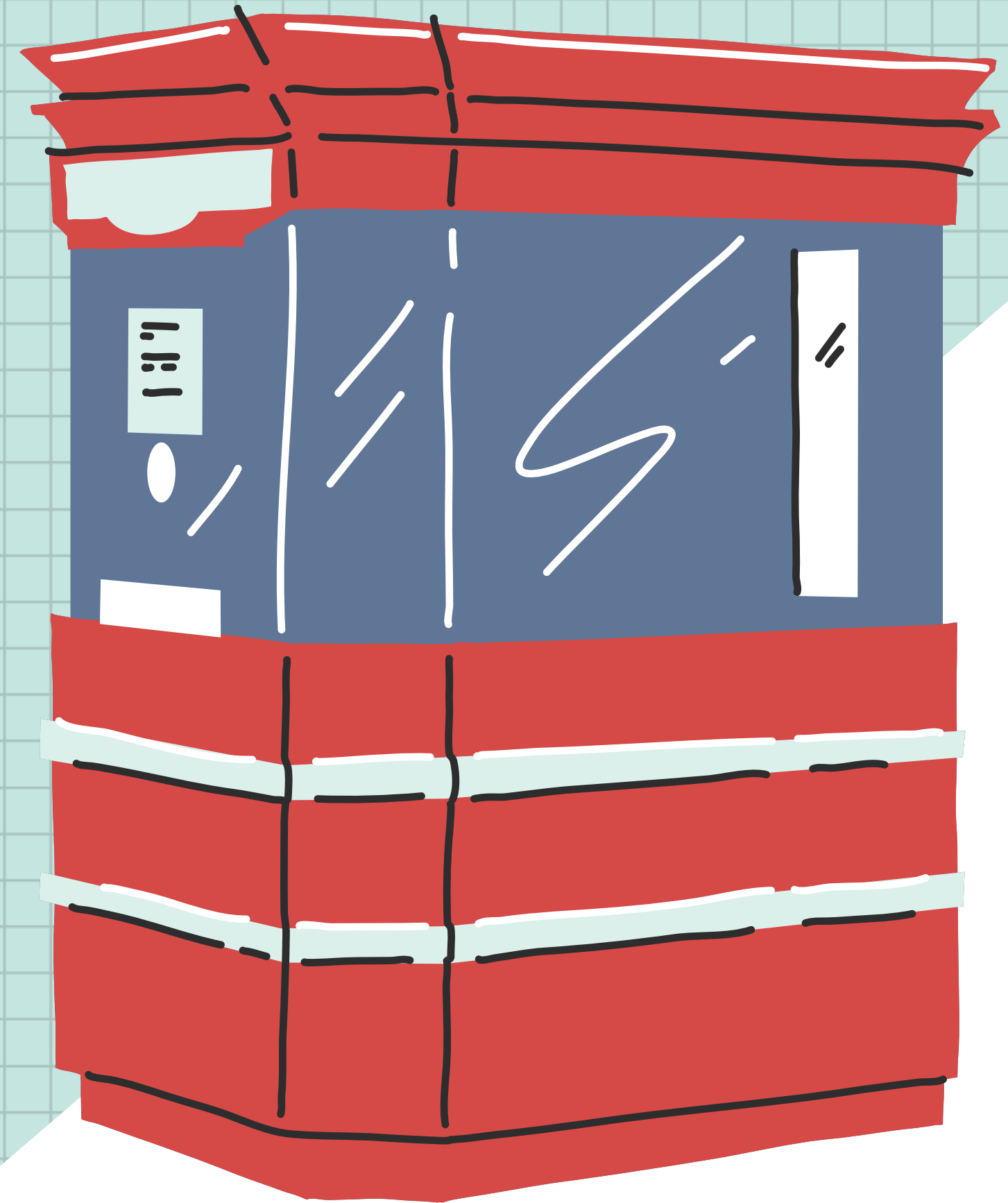
Reel Success

Film Profitability Prediction



Team Members

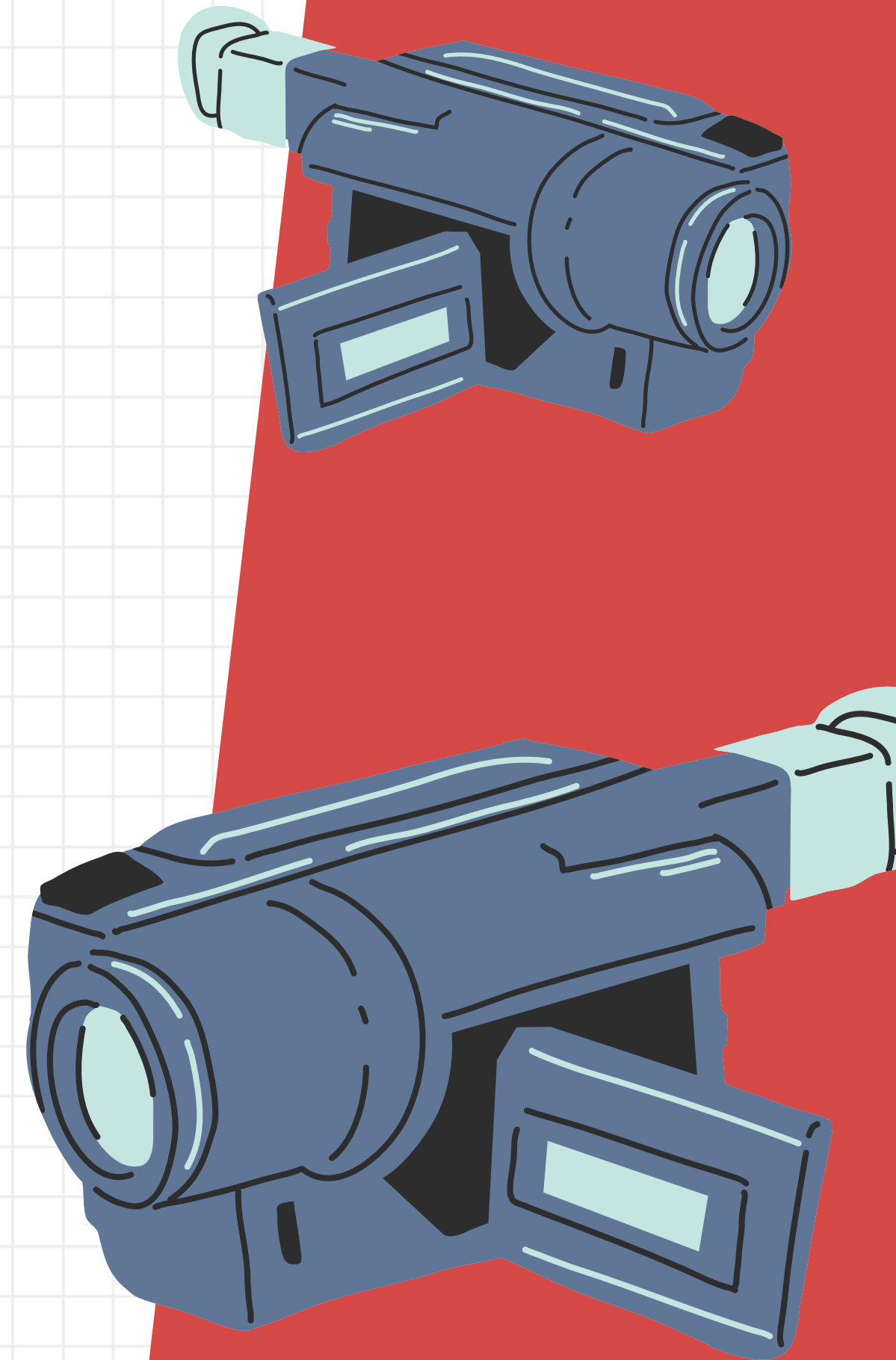
- 1 Angie Foust
- 2 Barb Rupps
- 3 Cassidy Simmons
- 4 Amanda Pigman



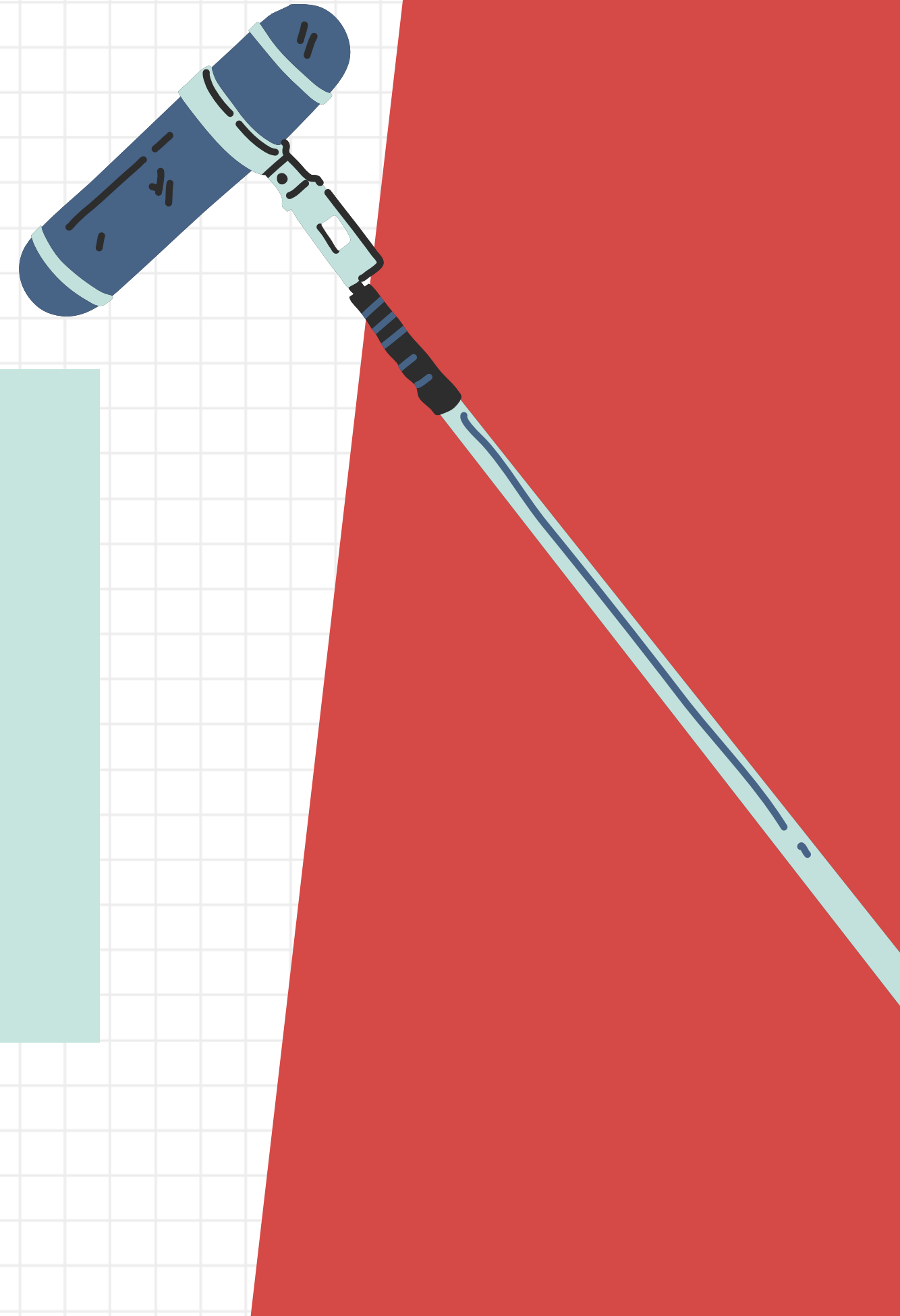
Overview

In this project, we used IMDB film data covering films for the years 1980-2020 from Kaggle to train and test a machine learning model that would predict the profitability of movies.

We used Pandas and Tableau to perform some exploratory data analysis to get an idea of things like outliers, correlations, and relationships between different features of the data.



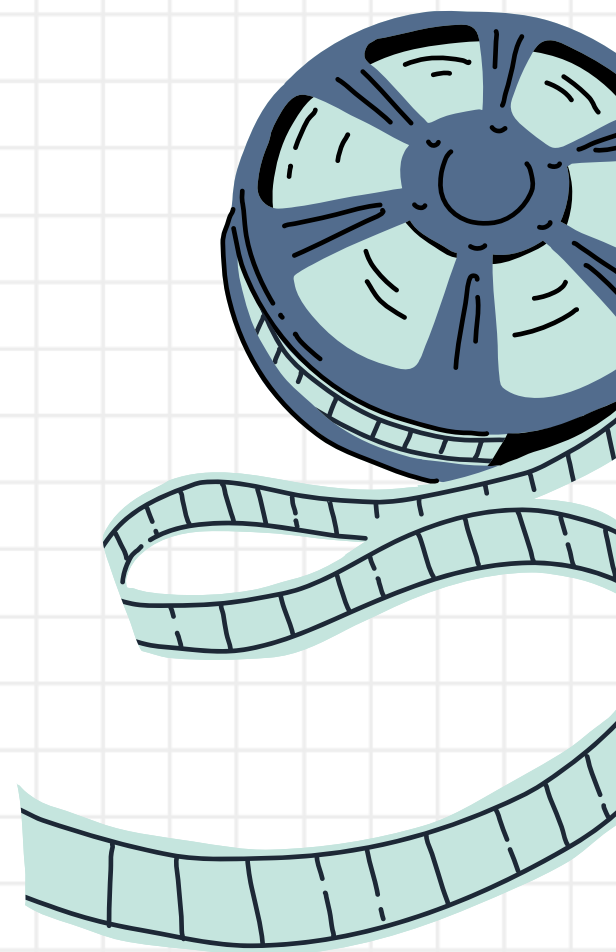
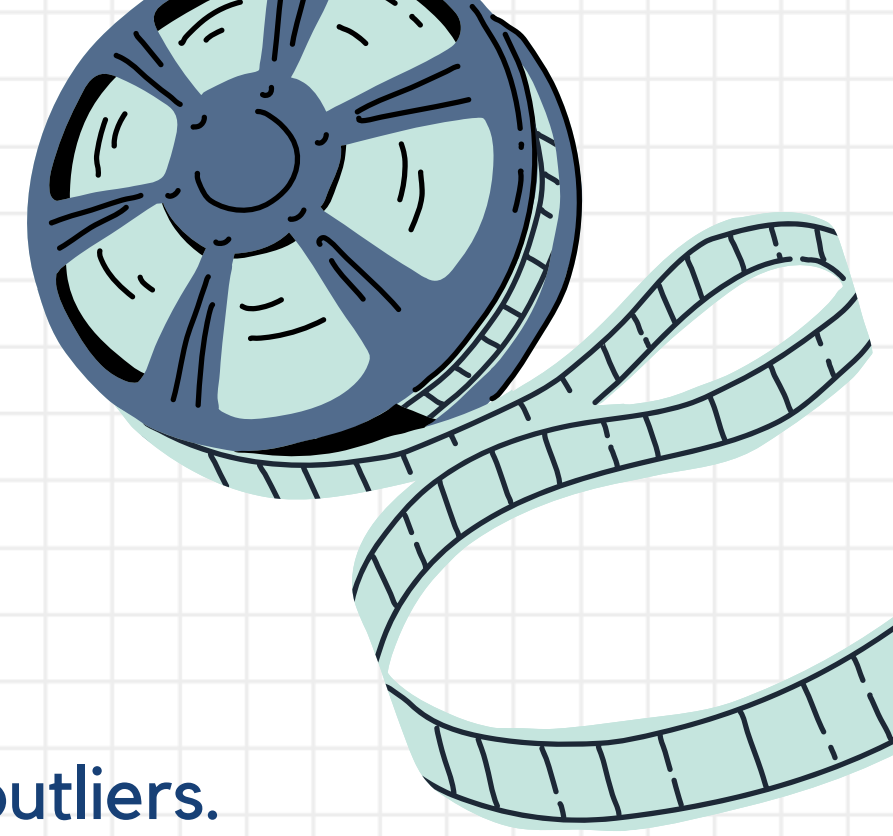
Defining Profitability



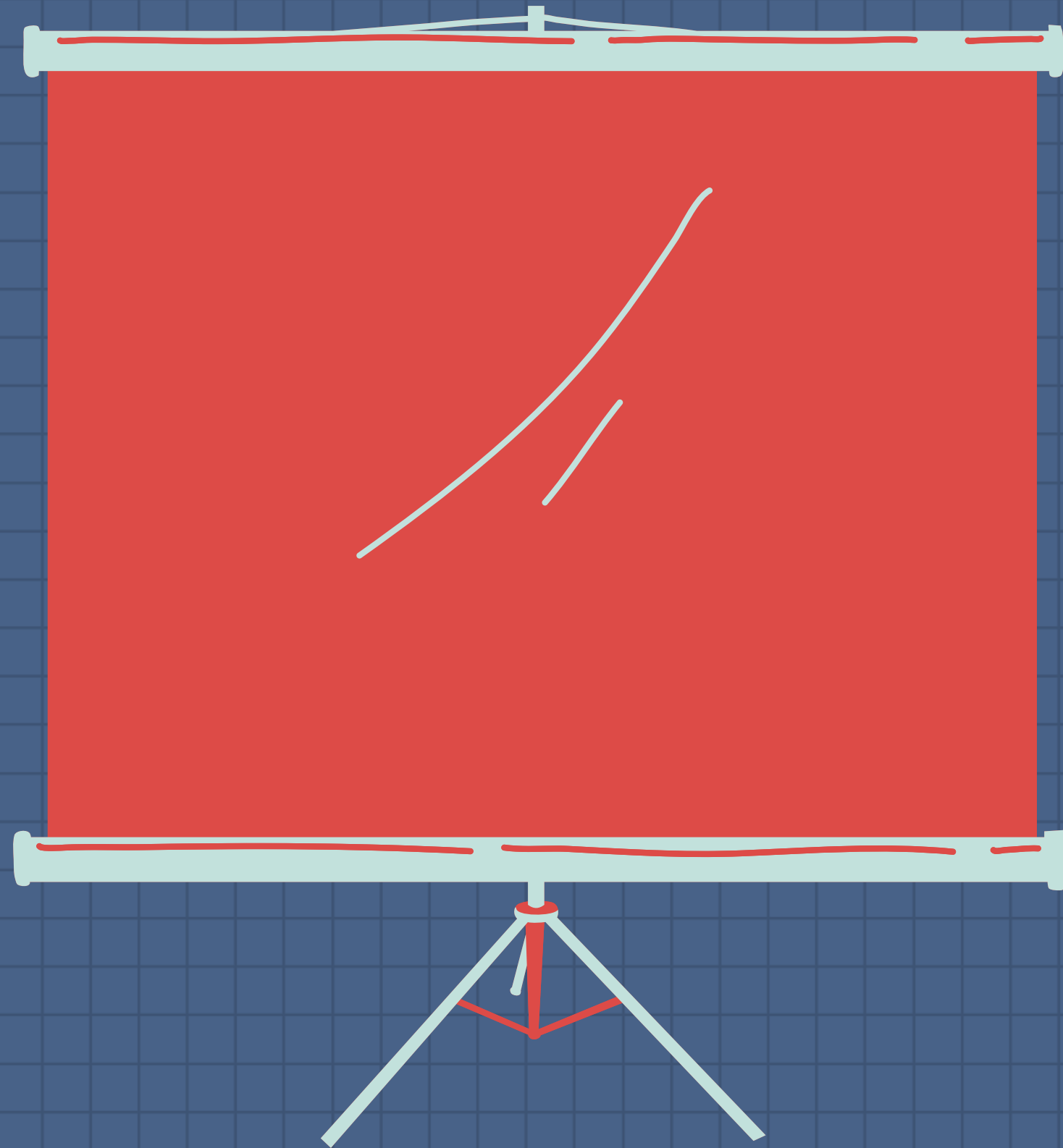
It is usually estimated that a movie needs to make 2-3 times its budget to become profitable in Hollywood after marketing costs, splitting revenue with theaters, and other costs are factored in.

Data

- Data source: Kaggle - movies from 1980 - 2020
- Data cleaning: used Spark to drop rows with null values & extreme outliers. Pandas was used to filter out features and other data cleaning functions.
- Data normalization and standardization: data was normalized and standardized using encoding and scaling techniques while building the model.



× ×



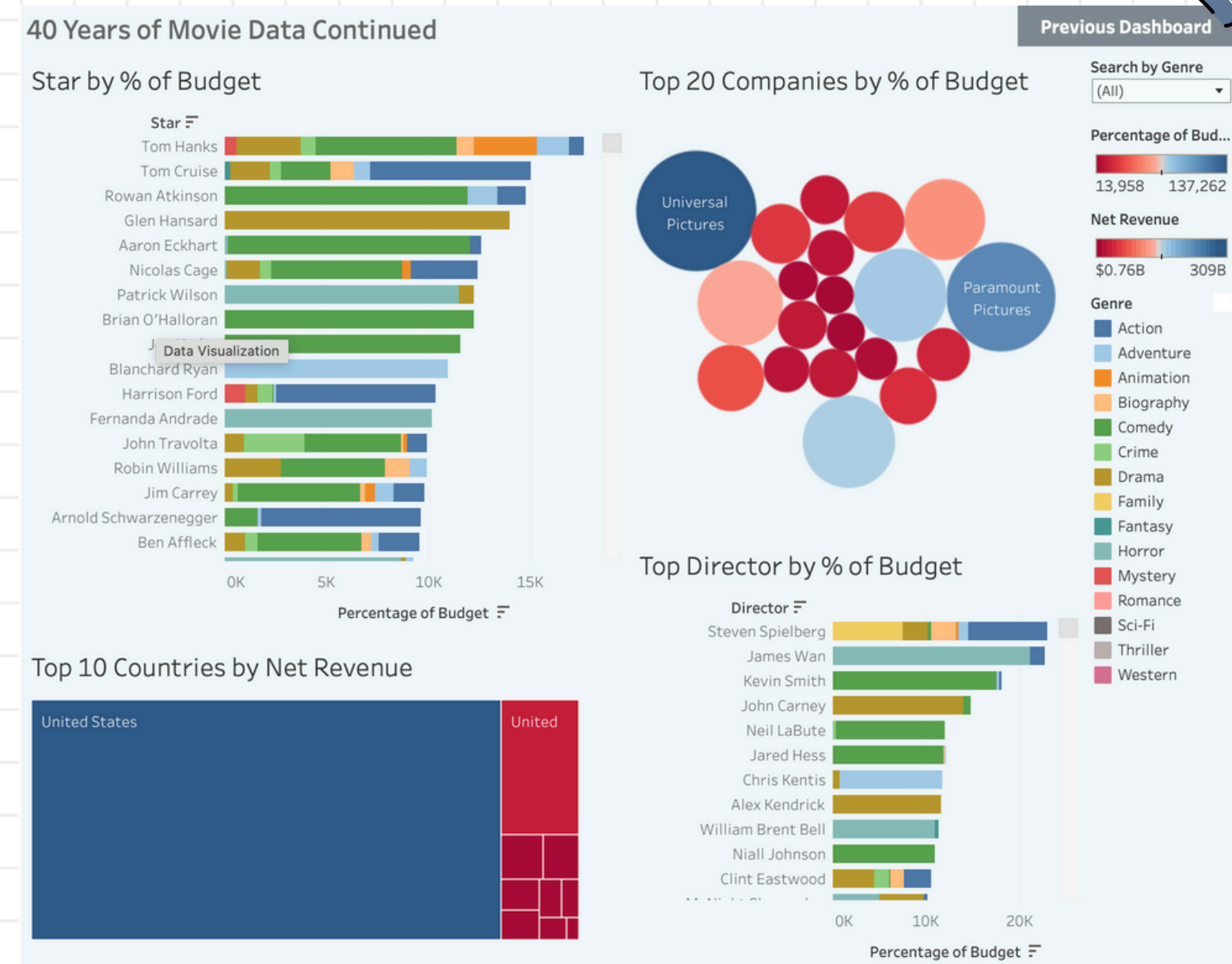
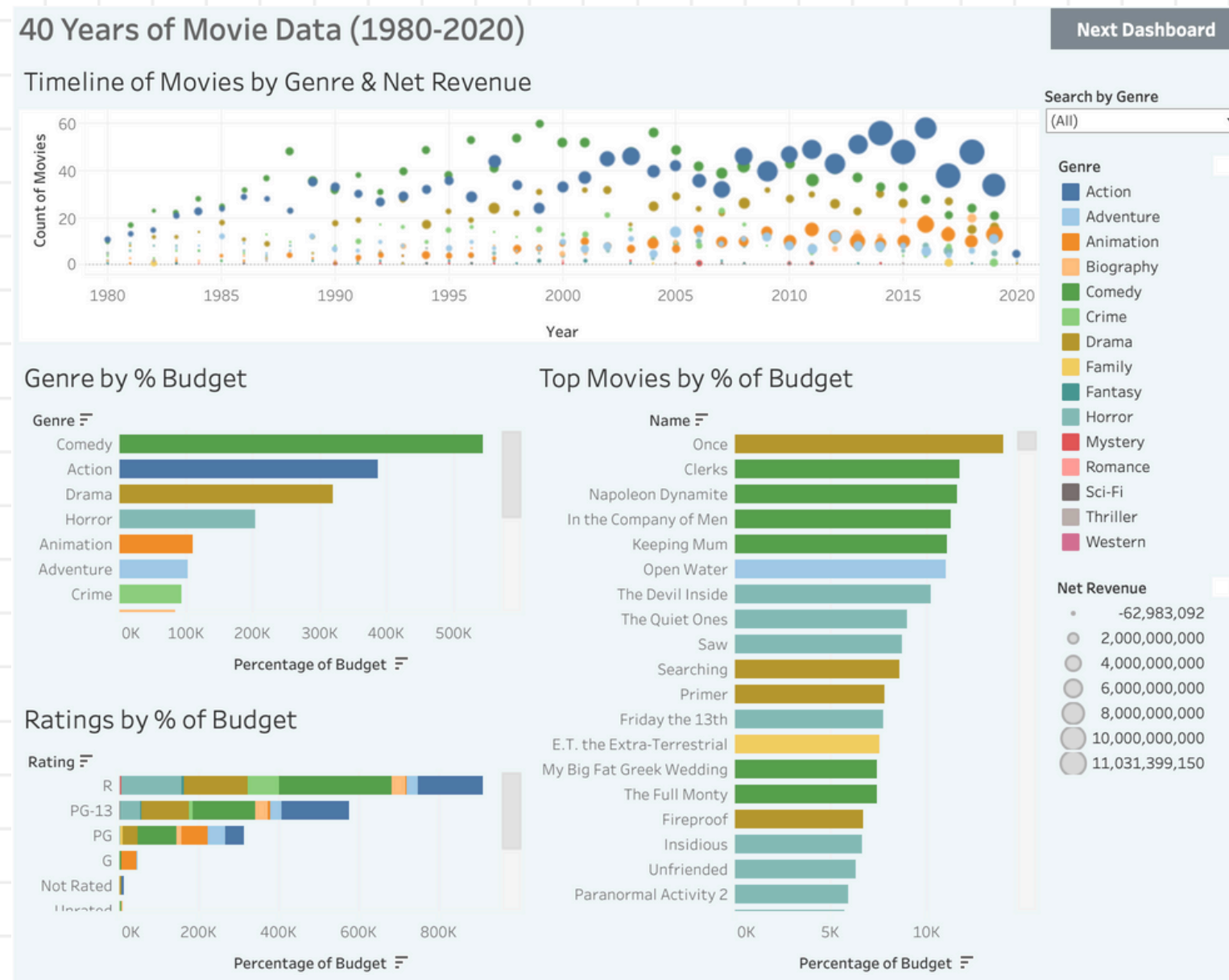
Visualizing
the data

× ×

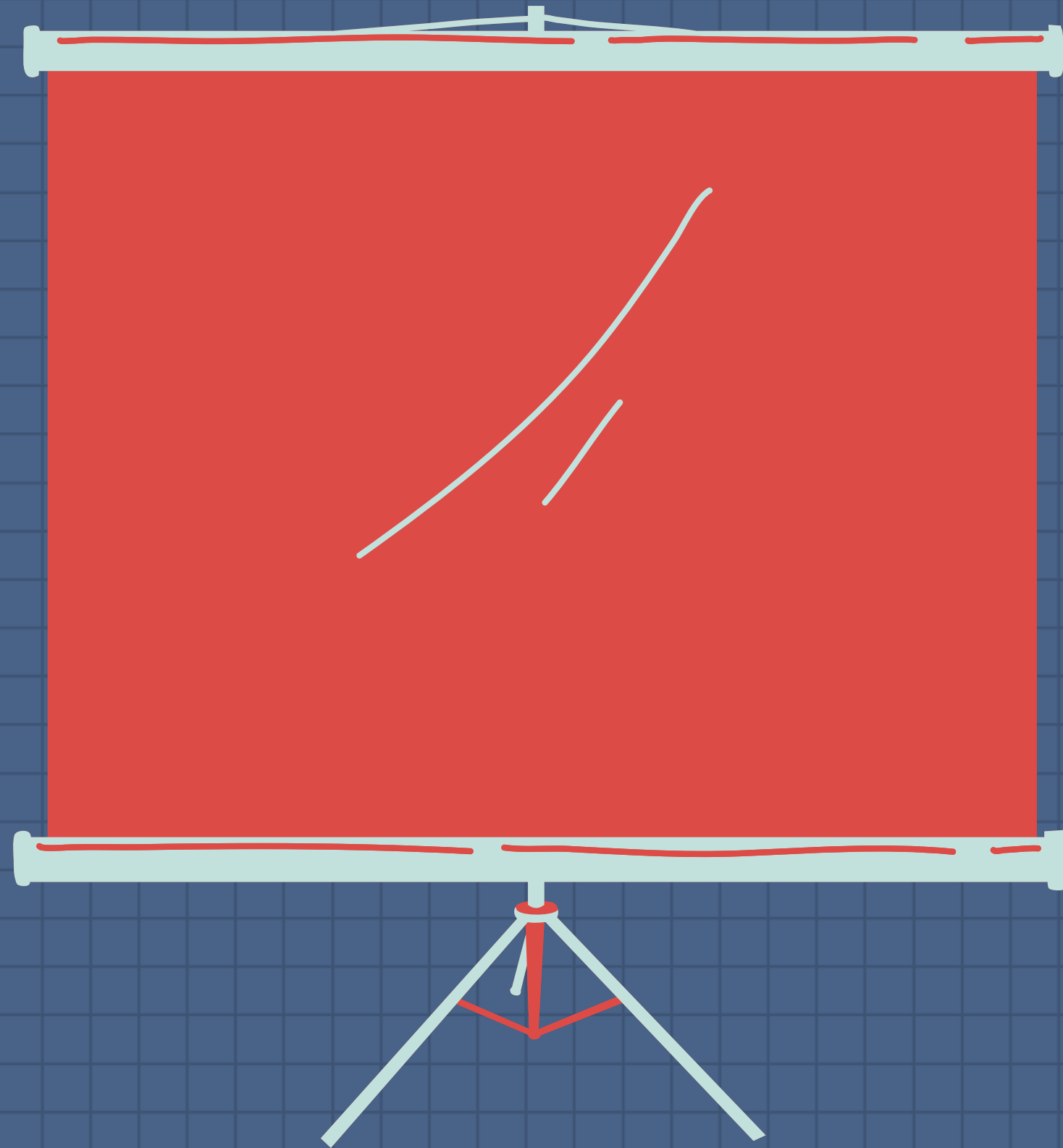
Tableau Visualizations



View the [Tableau dashboard](#) to browse the exploratory analysis we performed.

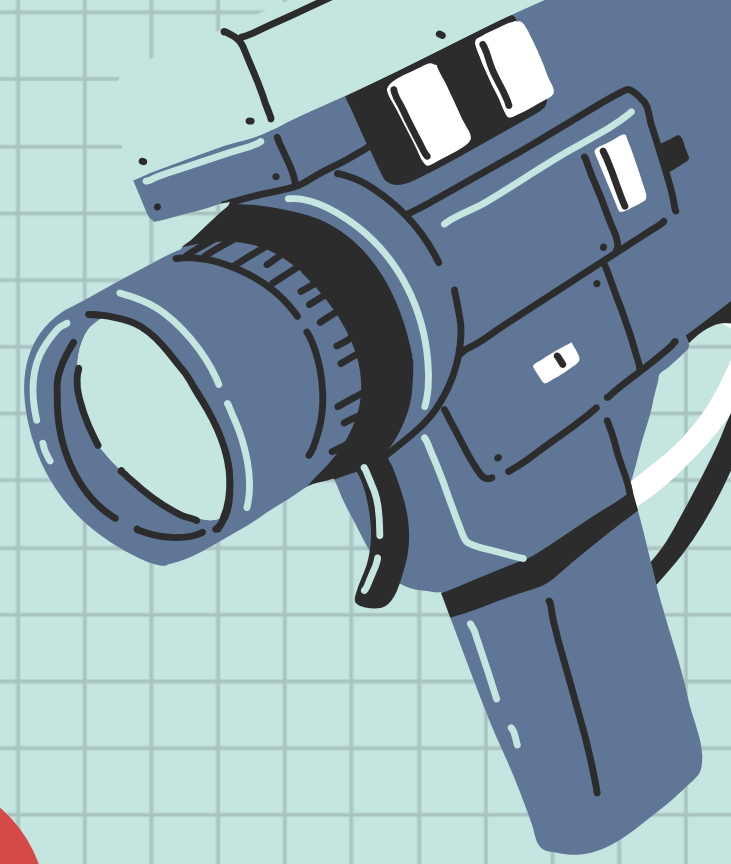
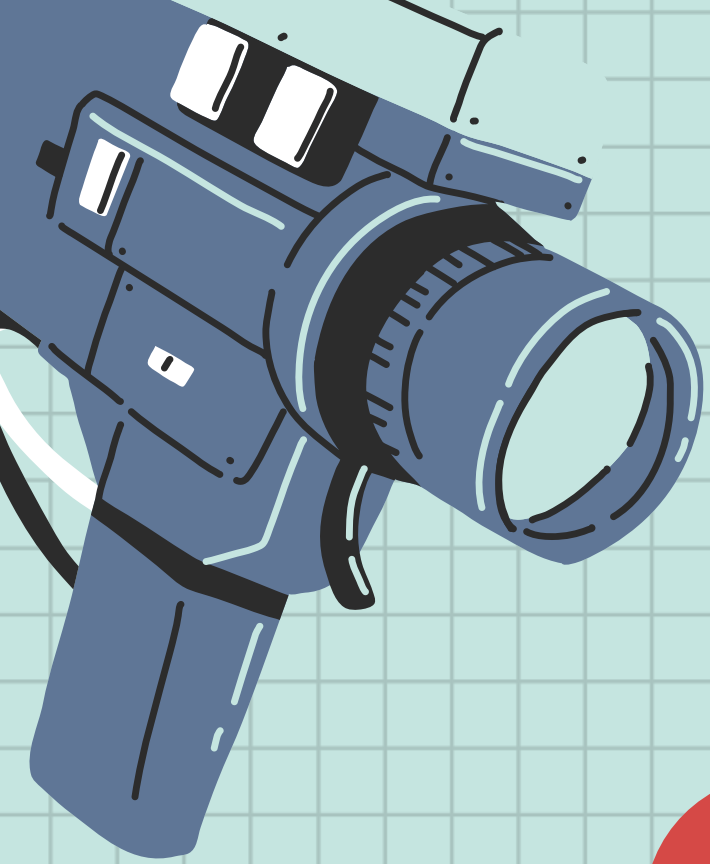


× ×



Choosing the
right Model

× ×



Models

Various combinations of features were tried resulting in model performances ranging from 30% to 65% accuracy.

1

Linear Regression

Performed on the two numerical columns that had the greatest correlation- budget vs. gross

2

Neural Network

Movies were marked either profitable or not profitable based on gross revenue divided by budget

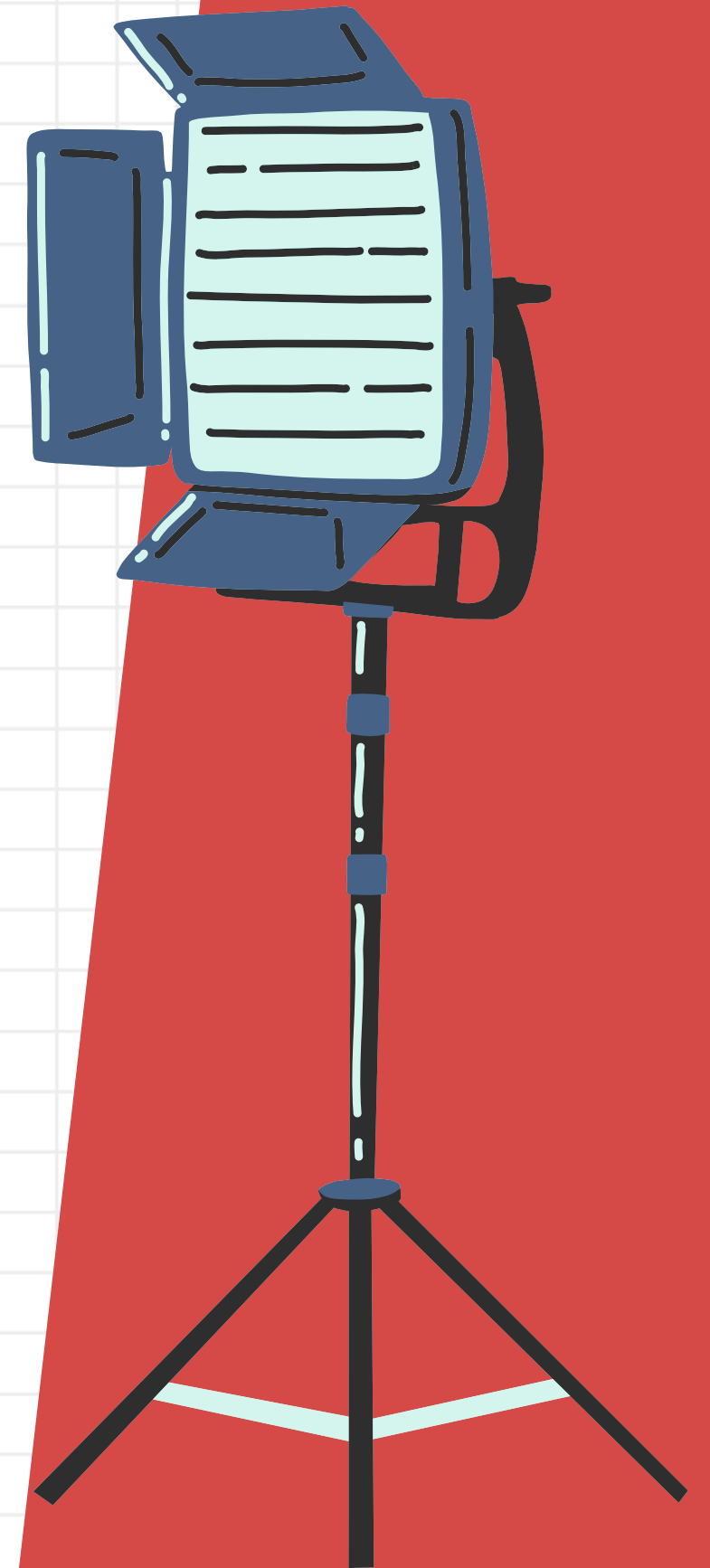
3

Random Forest

Again, movies were marked either profitable or not profitable based on gross revenue divided by budget

Considerations

After exploring the data further in Tableau, we determined there were four extreme outliers that were really skewing our data. We decided to remove them for that reason.



A stylized illustration of a film strip with light blue frames and dark blue borders, curving across a solid red background on the left side of the slide.

Conclusion

Our accuracy score of 65% is similar to accuracy scores achieved by other movie profitability prediction models.

Questions?

