

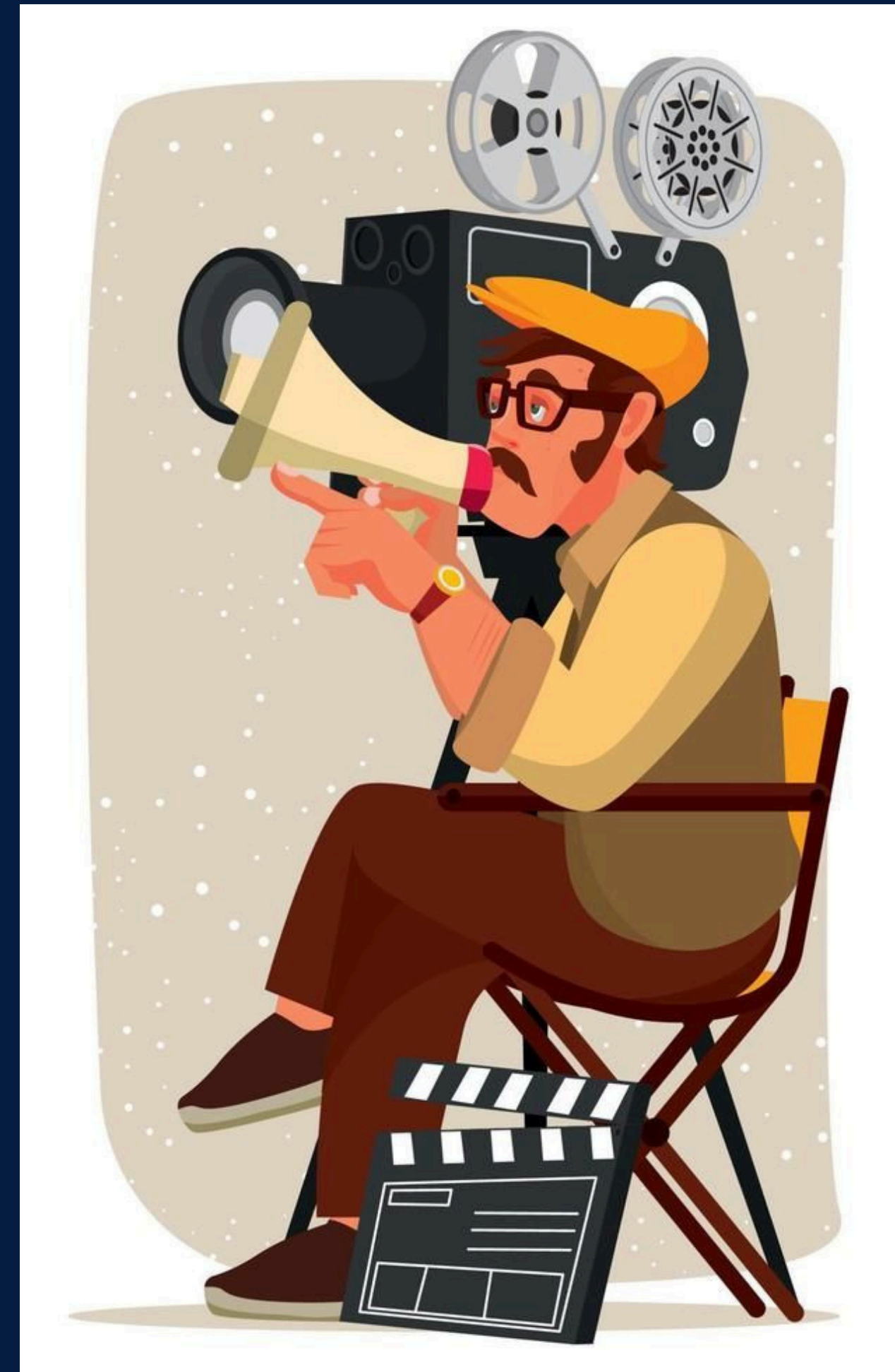
MOVIE RATING PREDICTION PROJECT

END-TO-END ML PROJECT |
PART OF CODSOFT DATA SCIENCE INTERNSHIP |
BY NISHANT SHARMA



PROBLEM STATEMENT

Can we predict a Movie's IMDb Rating using just its metadata - like based on Genre, Director and Actors before its release?



APPROACH TO SOLVE THIS PROJECT

DATA CLEANING & PREPARATION

LOADED THE MOVIE DATASET USING PANDAS AND IDENTIFIED MISSING/NULL VALUES USING STANDARD DATA INSPECTION TECHNIQUES.

CONVERTED ALL RELEVANT FIELDS INTO APPROPRIATE FORMATS — E.G., EXTRACTED NUMERICAL VALUES FROM STRINGS USING STRING OPERATIONS AND `PD.TO_NUMERIC()` FOR CLEAN NUMERICAL CONVERSION.

DATA VISUALIZATION & EDA

LEVERAGED SEABORN AND MATPLOTLIB TO GENERATE INSIGHTFUL VISUALIZATIONS:

HISTOGRAM PLOTS FOR IMDB RATINGS, MOVIE DURATION, RELEASE YEARS, AND NUMBER OF VOTES

HIGHLIGHTED TRENDS AND ANOMALIES TO GUIDE FEATURE SELECTION FEATURE ENGINEERING & MODEL PREP

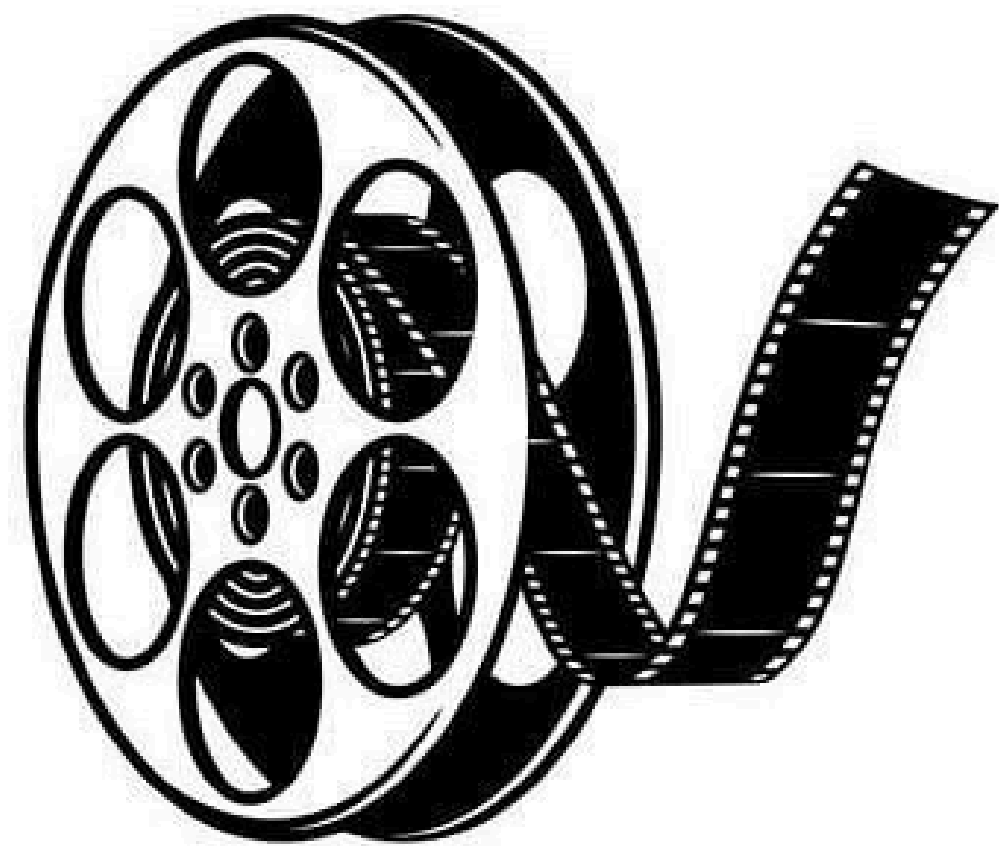
DEFINED TARGET VARIABLE (RATING) AND SELECTED FEATURE VARIABLES INCLUDING GENRE, DIRECTOR, CAST, YEAR, DURATION, AND VOTES.

SPLIT THE DATASET INTO TRAINING AND TESTING SETS TO ENABLE UNBIASED MODEL EVALUATION.

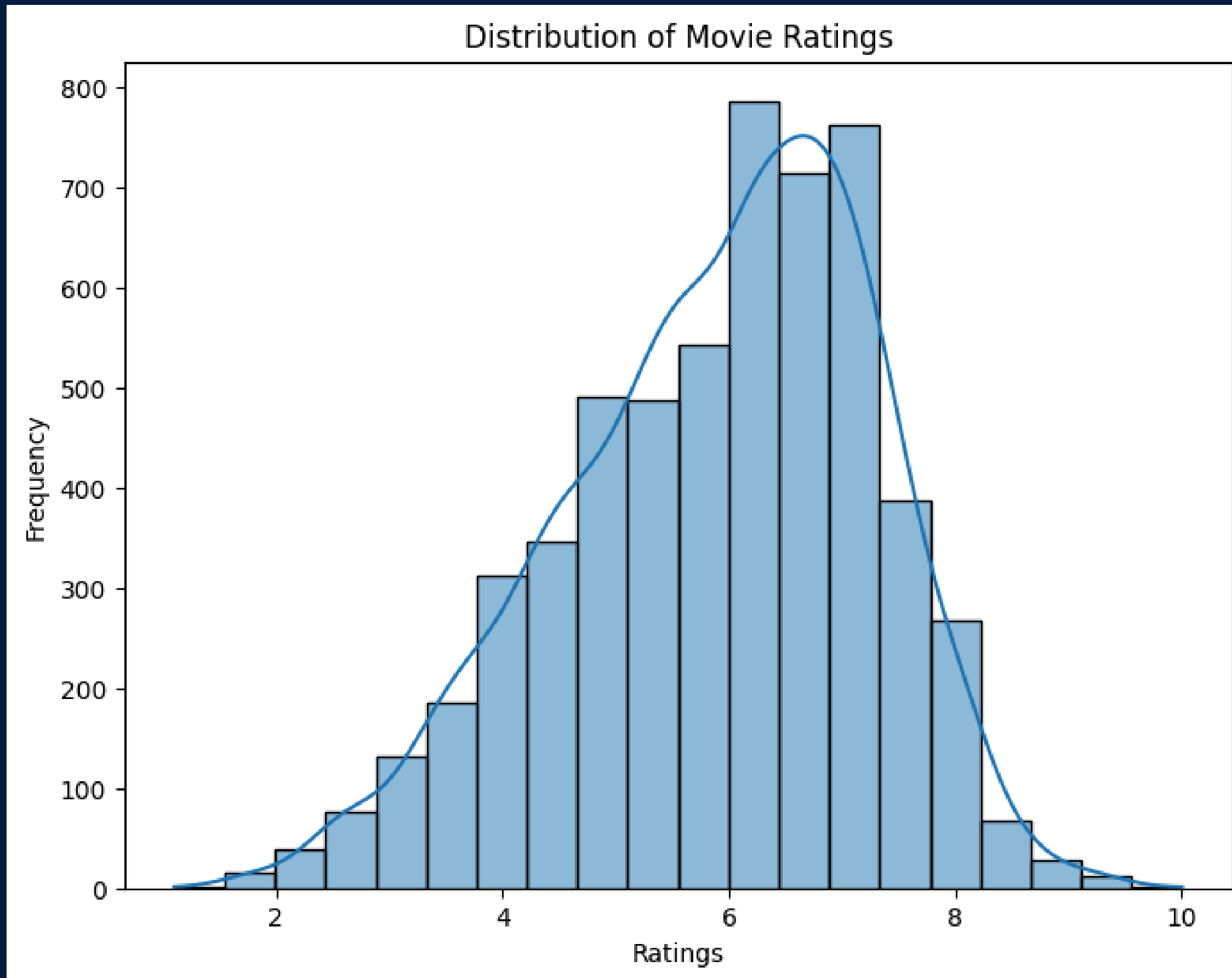
MODEL EVALUATION & ANALYSIS

TRAINED THE MODEL AND PLOTTED A SCATTER PLOT TO VISUALLY COMPARE ACTUAL VS. PREDICTED RATINGS, SHOWCASING PREDICTION ACCURACY.

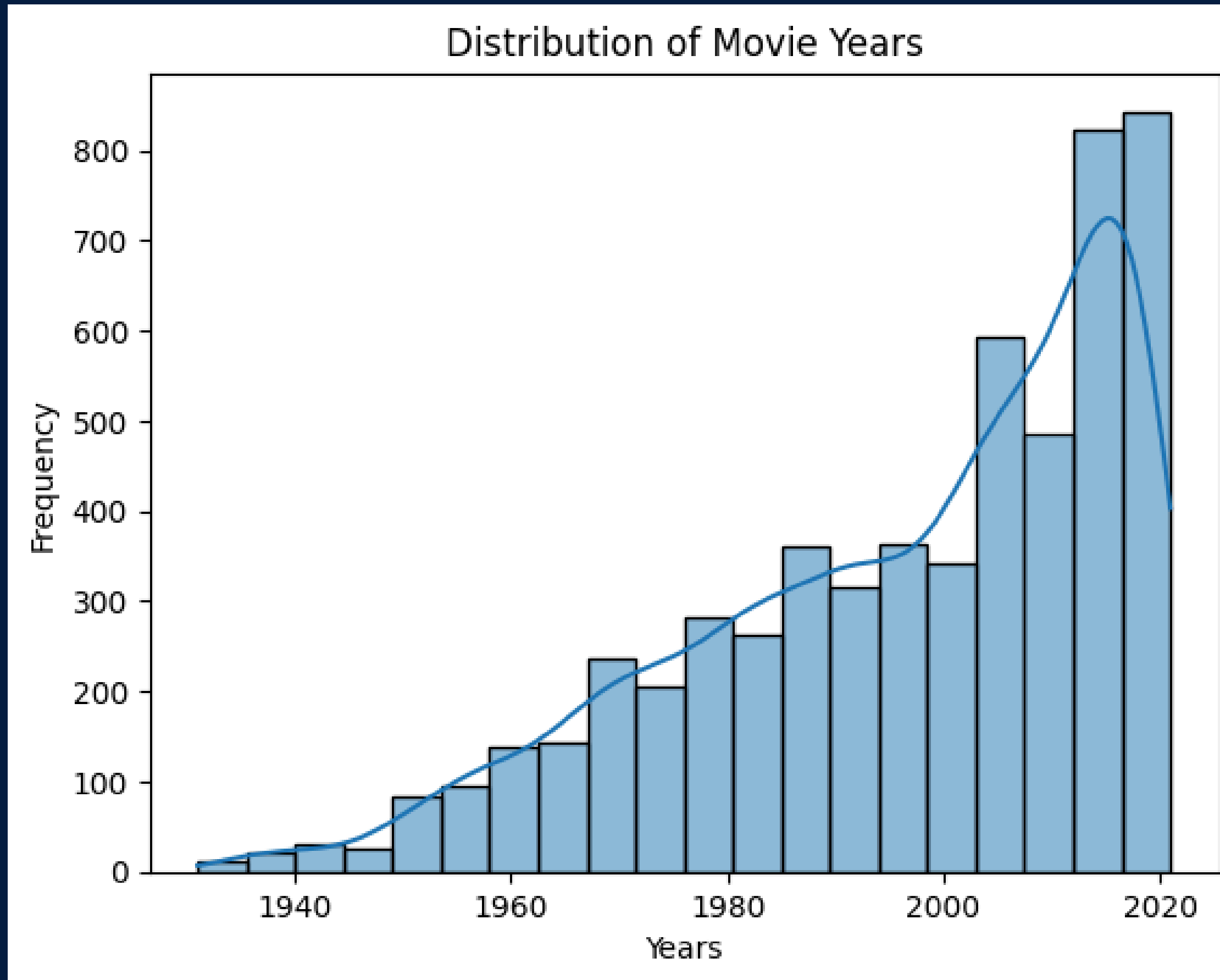
GENERATED A CORRELATION HEATMAP TO UNDERSTAND INTER-FEATURE RELATIONSHIPS AND MULTICOLLINEARITY.



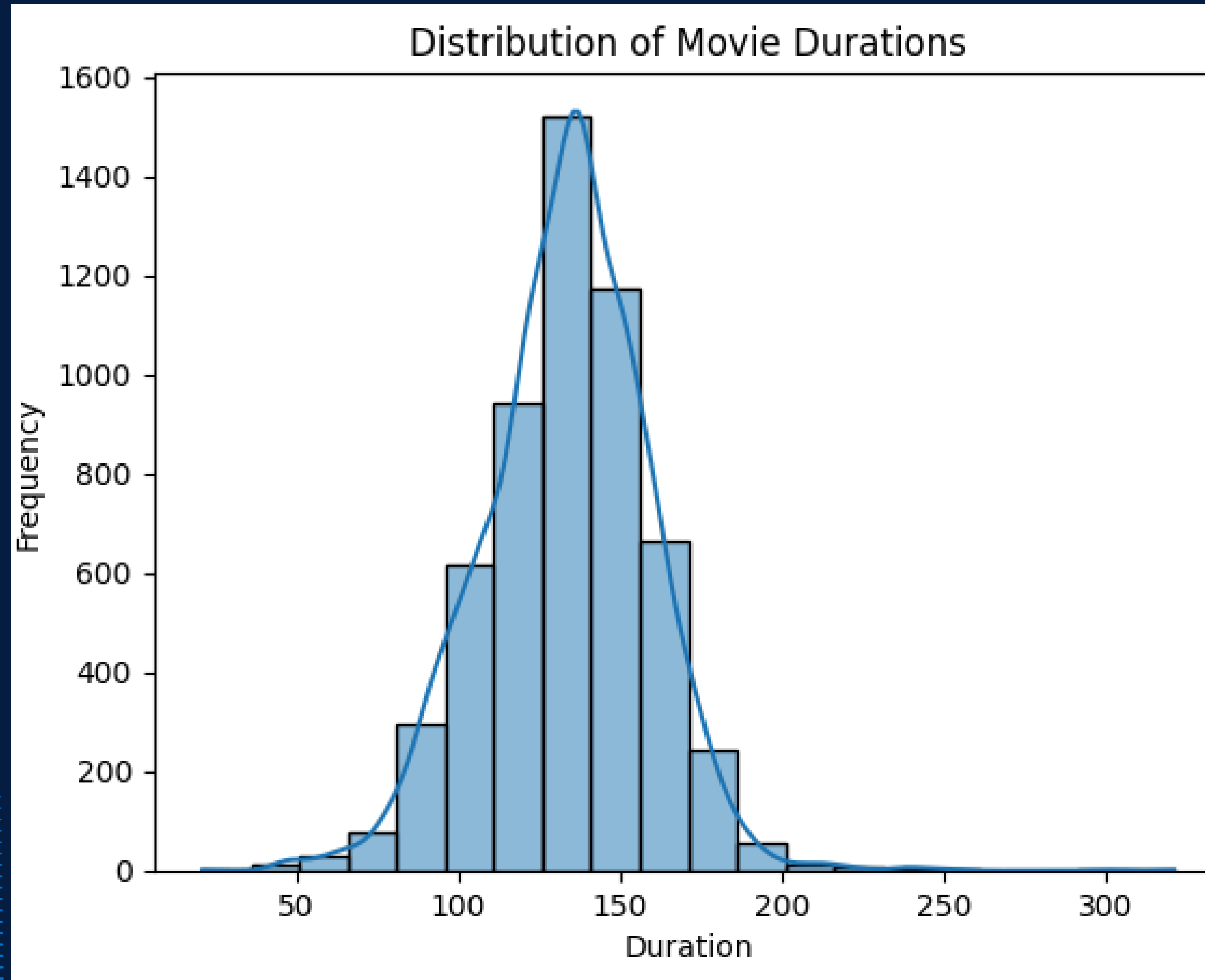
CHARTS AND VISUALIZATIONS



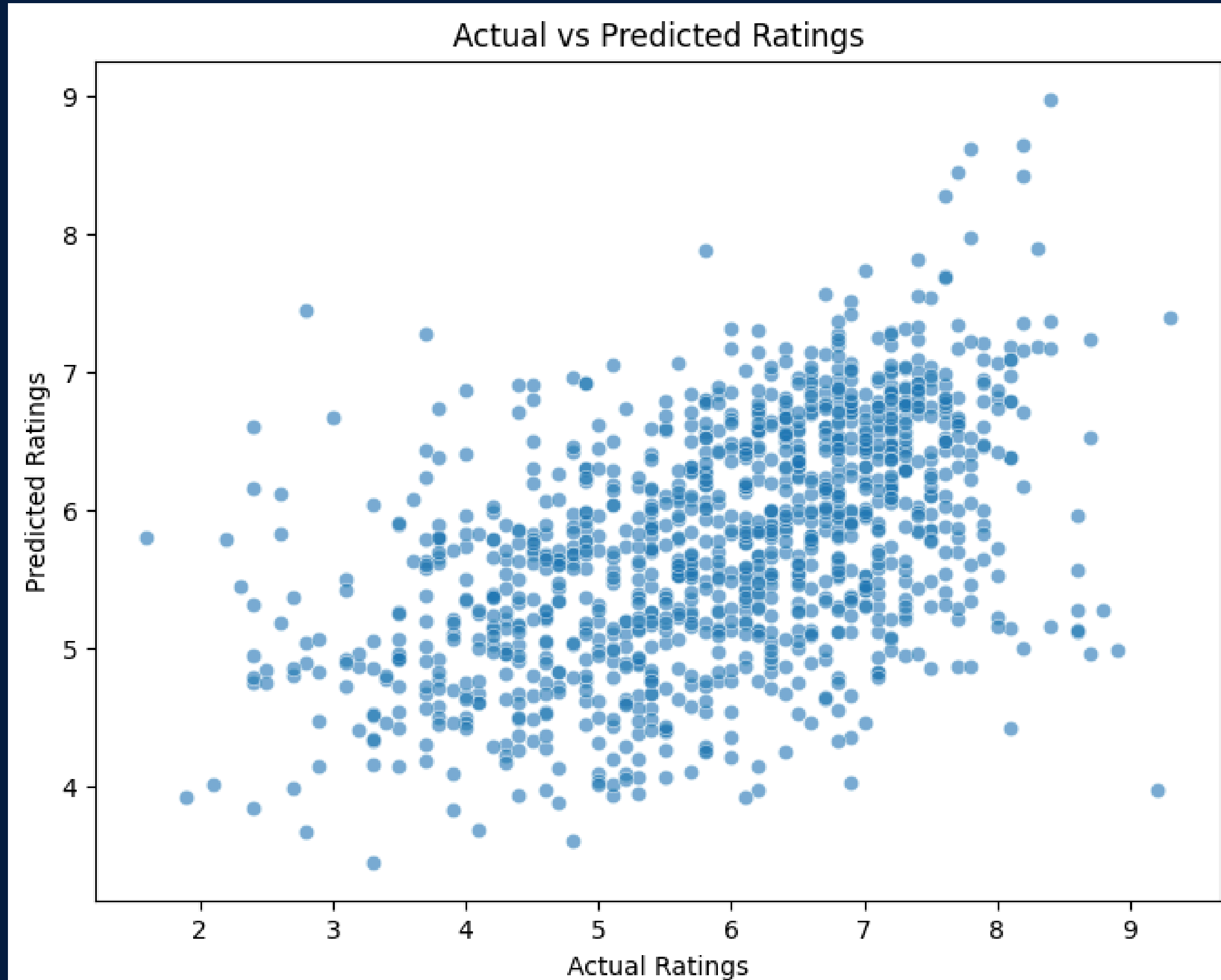
CHARTS AND VISUALIZATIONS



CHARTS AND VISUALIZATIONS

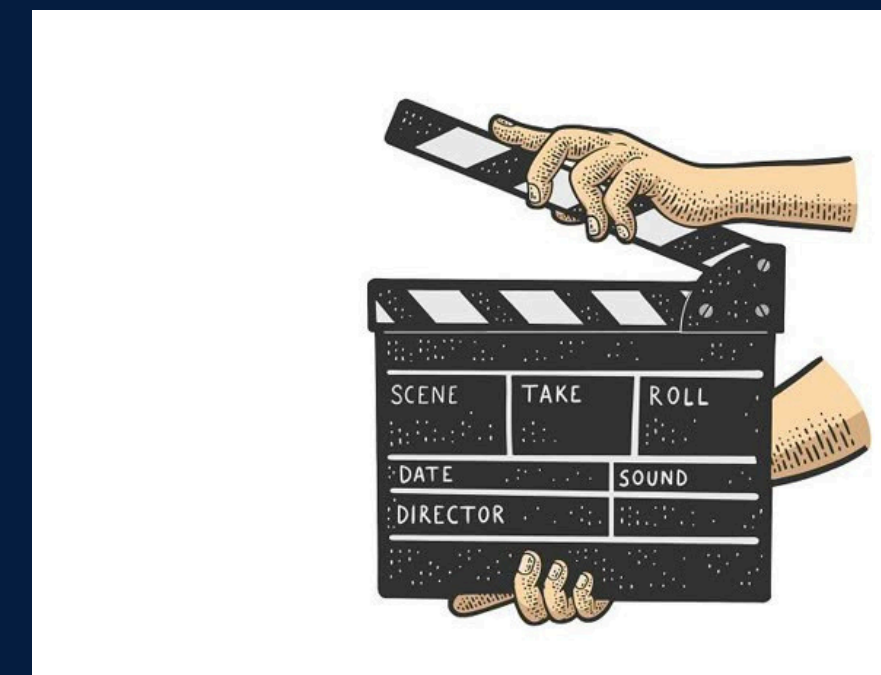
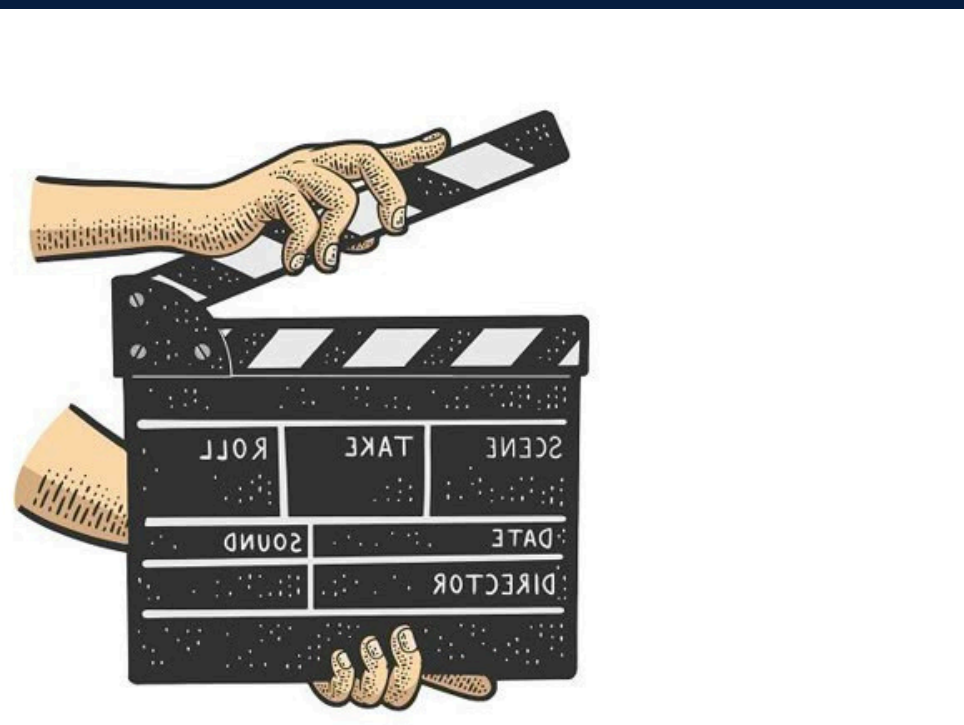


CHARTS AND VISUALIZATIONS



MY FINDINGS AND CONCLUSION

The linear regression model successfully predicts movie ratings based on three primary features: genre, director, and actors. The model achieved a rating of 6.96, demonstrating strong predictive performance. This indicates that these features are significant indicators of movie quality and audience reception.



The End