Movie Earnings

Kurt Brown, Matt Latzke, & Nick Borchich

Purpose

- Better understand modern culture’s values
- Find out what stories appeal to people
- Determine what people want from entertainment today
- Find a way to become moneymakers by finding what people like most

Design

Methods we used:

- Histograms
- Heatmaps
- r, R^2, adjusted R^2, and RSME
- T-tests
- Averages
- Linear regression
- Decision trees
- Random forest classifications
- Confusion matrices

Achievements

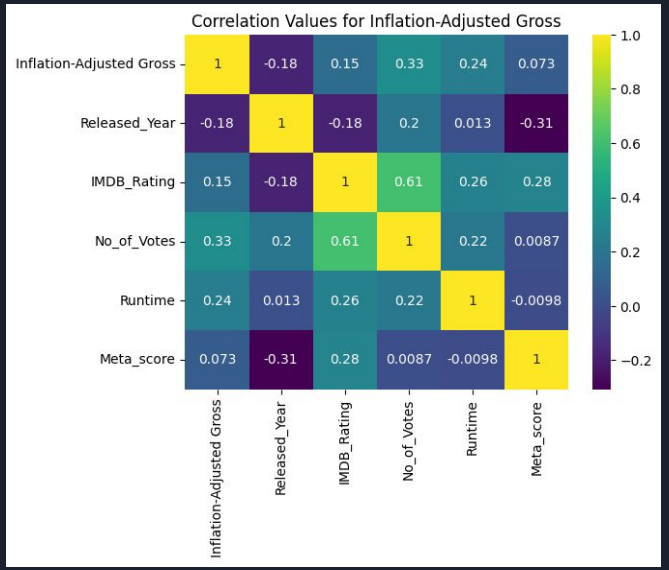
We managed to:

- Meet very minimally outside of class
- Find mostly insignificant results
- Not procrastinate too much

We learned that:

- Editing the same google colaboratory simultaneously is problematic
- You can find insignificant results and still get good grades
- Some models are a pain to use with over 20 variables

Initial Results (Heatmap & R^2 Values)



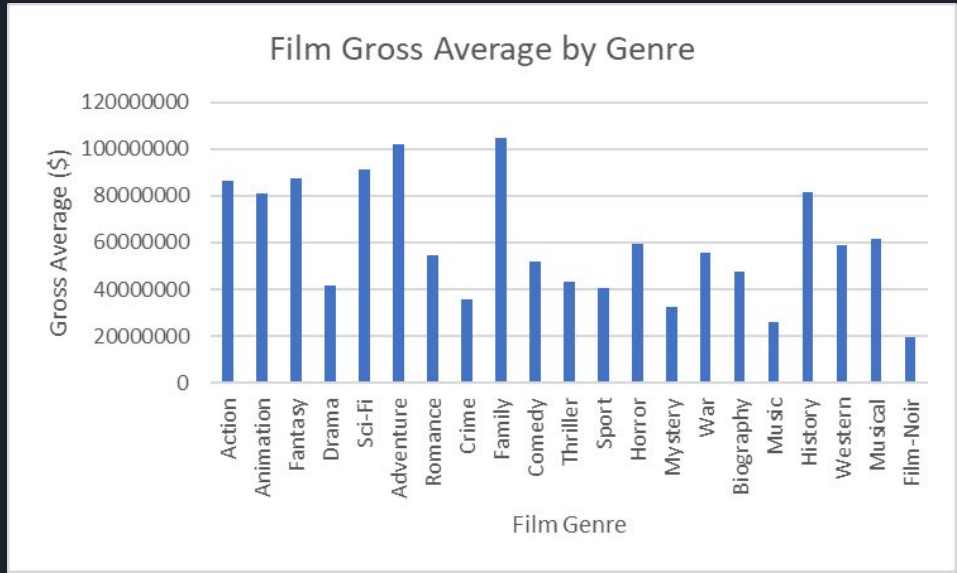
We used the above correlation heatmap to find out which of our explanatory variables were most correlated with inflation-adjusted gross.

We also did some calculations and found the R^2 values for each of our explanatory variables. After viewing the results, it is very sad to see that the value .106 relating to the number of votes was our highest R^2 value.

	Release Year vs Gross	Release Year vs Adjusted Gross	IMDB Rating vs Adjusted Gross	Number of Votes vs Adjusted Gross	Runtime vs Adjusted Gross	Runtime vs Adjusted Gross
R^2 value	0.055	0.031	0.023	0.106	0.057	0.0052

Genre Average Results

In order to try and get a gauge for how genre was as a predictor for inflation-adjusted gross, we figured we could take the averages of each genre in order to see if certain genres averaged significantly more. We certainly did find that genres like family and adventure earned significantly more on average than music and film-noir movies.

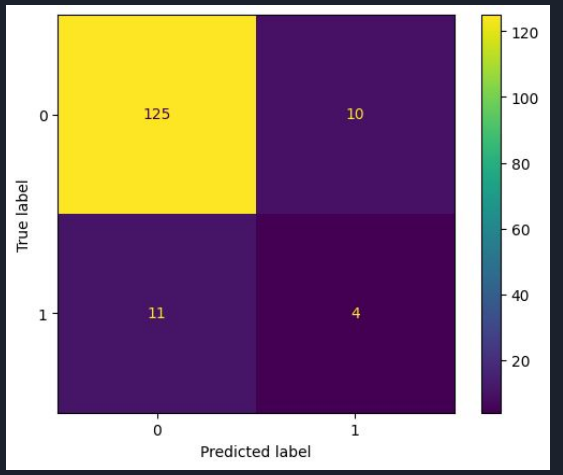


Regression Model Results

- All were pretty inaccurate
- Best model used number of votes, IMDB rating, our dummy variable for famous director or not, and another dummy variable for determining if the movie's genre was either action or adventure
- R^2=0.179 with a Root Mean Squared Error of 61,700,609.54

Decision Tree & Random Forest Predictions

Above we have our decision tree model results. We were able to achieve an F_1 score of 0.6 while also receiving the exact same score for precision and recall.



For our random forest model results on the same data, we were able to slightly increase our F_1 score to .68 with a 0.66 precision value and 0.69 recall value.