

## LINEAR REGRESSION

### 1. Predict IMDb score based on Runtime

The expected result from a linear regression model that predicts IMDb score based on runtime would be a regression equation that describes the linear relationship between the two variables. Specifically, the equation will give you an estimate of the intercept (the IMDb score when the runtime is zero) and the slope (the change in IMDb score for every one-unit increase in runtime). The equation will look like:

$$\text{IMDb score} = \text{Intercept} + \text{Slope} * \text{Runtime}$$

The regression output will also include other statistics such as the R-squared value, which indicates how much of the variability in the IMDb scores can be explained by the runtime, and the p-value for the slope, which indicates whether the relationship between runtime and IMDb score is statistically significant.

The R-squared value of 0.0028 indicates that only 0.28% of the variance in IMDb score is explained by the runtime variable alone. This suggests that there may be other factors that influence the IMDb score of a movie or TV show beyond just its runtime.

### 2. Predict IMDb votes based on Release year

The results of the linear regression analysis suggest a weak positive relationship between release year and IMDb votes, with more recent releases tending to receive slightly higher IMDb votes. However, the low R-squared value indicates that the release year variable alone is not a strong predictor of IMDb votes, and other factors such as the quality of the plot, the cast, the production value, or the marketing efforts may also play a significant role in determining the number of IMDb votes.

$$\text{IMDb votes} = 502511 - 2480.6 * \text{release year}$$

This equation suggests that the IMDb votes increase by 2480.6 for every one-year increase in release year. The intercept of -502511 indicates the number of IMDb votes for a movie or TV show released in the year 0, which is not possible, but it is included in the model for mathematical purposes.

The R-squared value of 0.0032 indicates that only 0.32% of the variance in IMDb votes is explained by the release year variable alone. This suggests that there may be other factors that influence the number of IMDb votes a movie or TV show receives beyond just its release year.

## MULTIPLE REGRESSION

### 1. Predicting TMDB score based on TMDB popularity and Runtime

The multiple regression analysis showed that there was a significant relationship between TMDB score, TMDB popularity, and runtime. The R-squared value of the model was 0.053, indicating that 5.3% of the variation in TMDB score could be explained by the independent variables of TMDB popularity and runtime. The p-values for the coefficients of TMDB popularity and runtime were less than 0.05, indicating that both variables were statistically

significant in predicting TMDB score. Thus, we can conclude that TMDB popularity and runtime are useful predictors of TMDB score.

TMDB score is a measure of the popularity of a movie or TV show on the TMDB platform, while TMDB popularity measures the relative popularity of a title among TMDB users. Runtime, on the other hand, is the length of the movie or TV show in minutes. By analyzing the relationship between these variables, we can determine the extent to which TMDB popularity and runtime can be used to predict TMDB score.