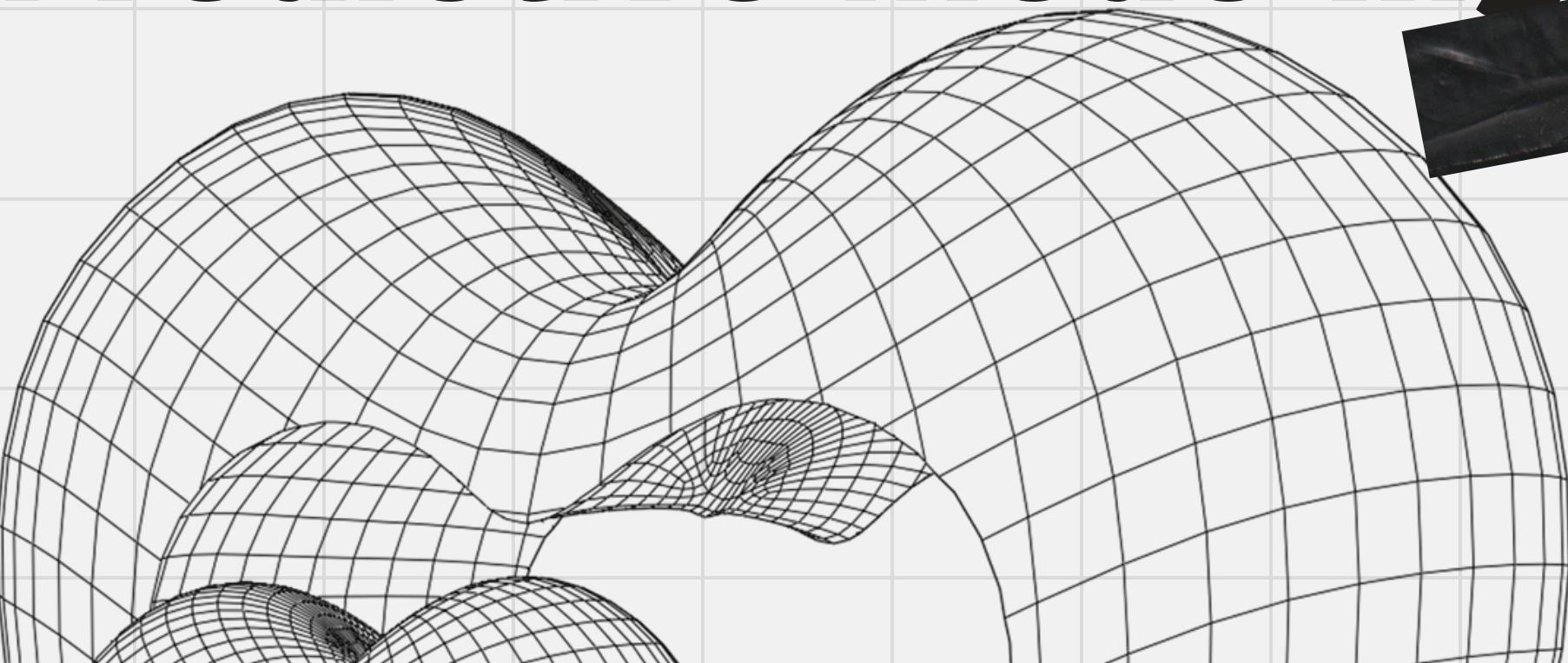


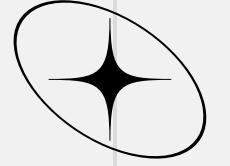
CARREON | LUCAS | PERALTA | QUIPIT

Optimizing Ad Bidding Strategies: Analyzing Facebook's A/B Test with Predictive Modeling



GROUP 4



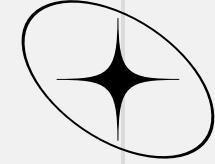


INTRODUCTION

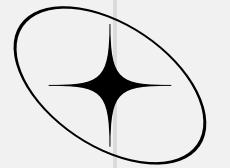
In the competitive world of digital advertising, this analysis leverages an A/B test dataset to evaluate the effectiveness of various Facebook ad strategies, focusing on key metrics such as impressions, clicks, purchases, and earnings. We use machine learning models like Linear Regression, Decision Tree Regressor, and Gradient Boosted Trees (GBT) Regressor to predict log-transformed earnings based on these metrics, employing feature engineering techniques to enhance performance. Evaluation through Root Mean Squared Error (RMSE) and R-squared (R²), along with visualizations of model predictions and residuals, reveals insights into each model's accuracy and areas for improvement. These findings provide actionable insights to guide strategic decisions for optimizing future ad campaigns and maximizing ROI.



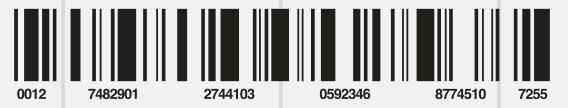
DATA DICTIONARY

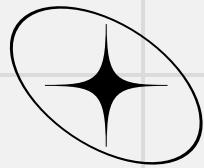


Attribute	Description
Impression	Number of times an advertisement was displayed to users.
Click	Number of times users clicked on an advertisement.
Purchase	Number of purchases made after clicking on an advertisement.
Earning	Revenue generated from purchases.
Group	Category indicating whether the data belongs to the control group (Control) or the test group (Test).
features	A vector containing the Impression, Click, and Purchase attributes, used as input features for machine learning models.
poly_features	A vector containing polynomial features generated from the features vector.
scaled_features	A vector containing scaled polynomial features, ensuring that all features have a similar range.
log_Earning	Log-transformed earnings to handle skewness in the Earning attribute.
prediction	Predicted log earnings from the machine learning models.



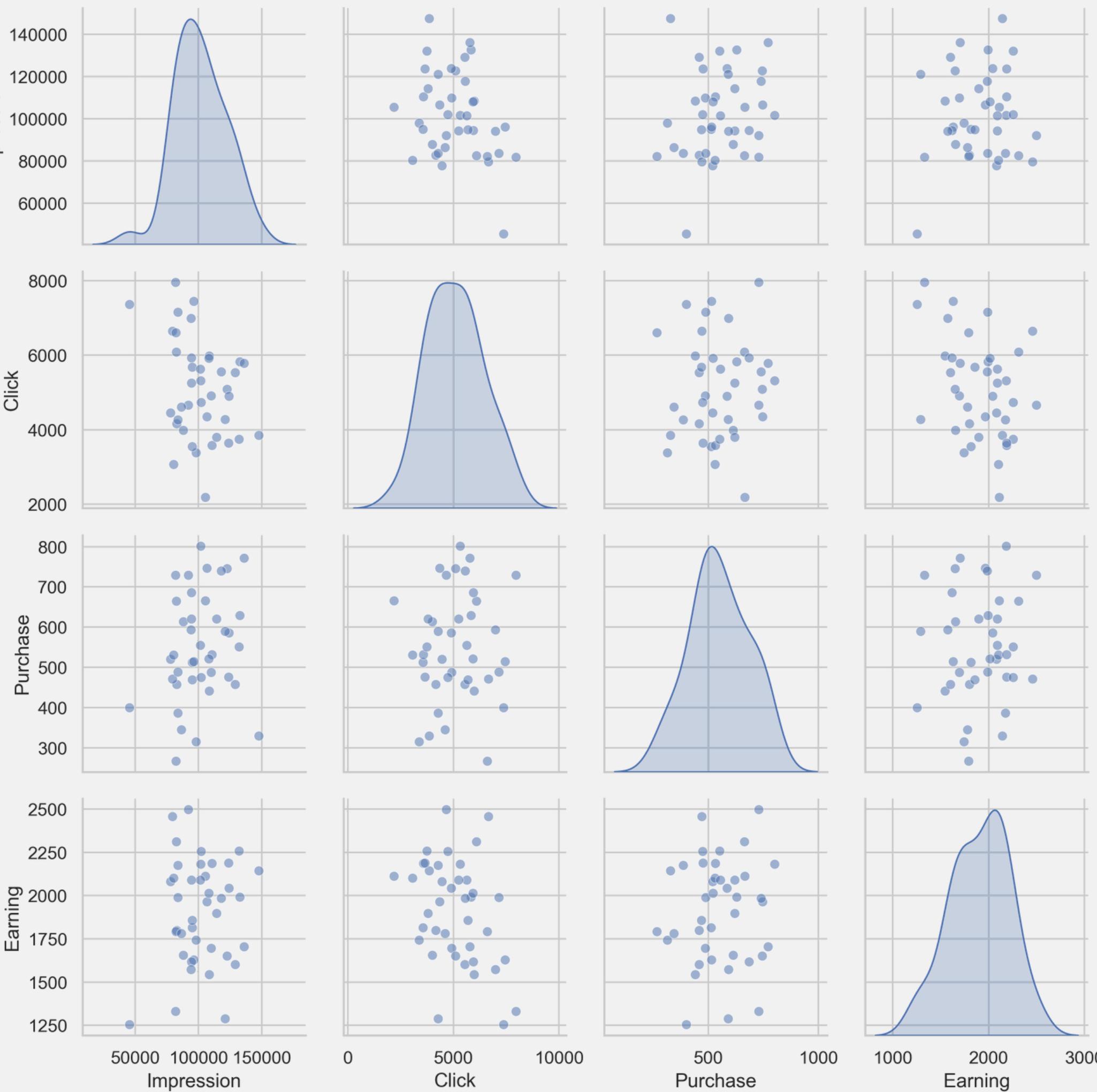
WHICH MACHINE LEARNING MODEL BEST PREDICTS AD CAMPAIGN EARNINGS USING IMPRESSIONS, CLICKS, AND PURCHASES?

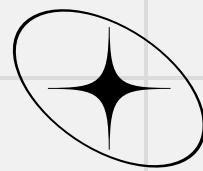




CONTROL GROUP DISTRIBUTIONS

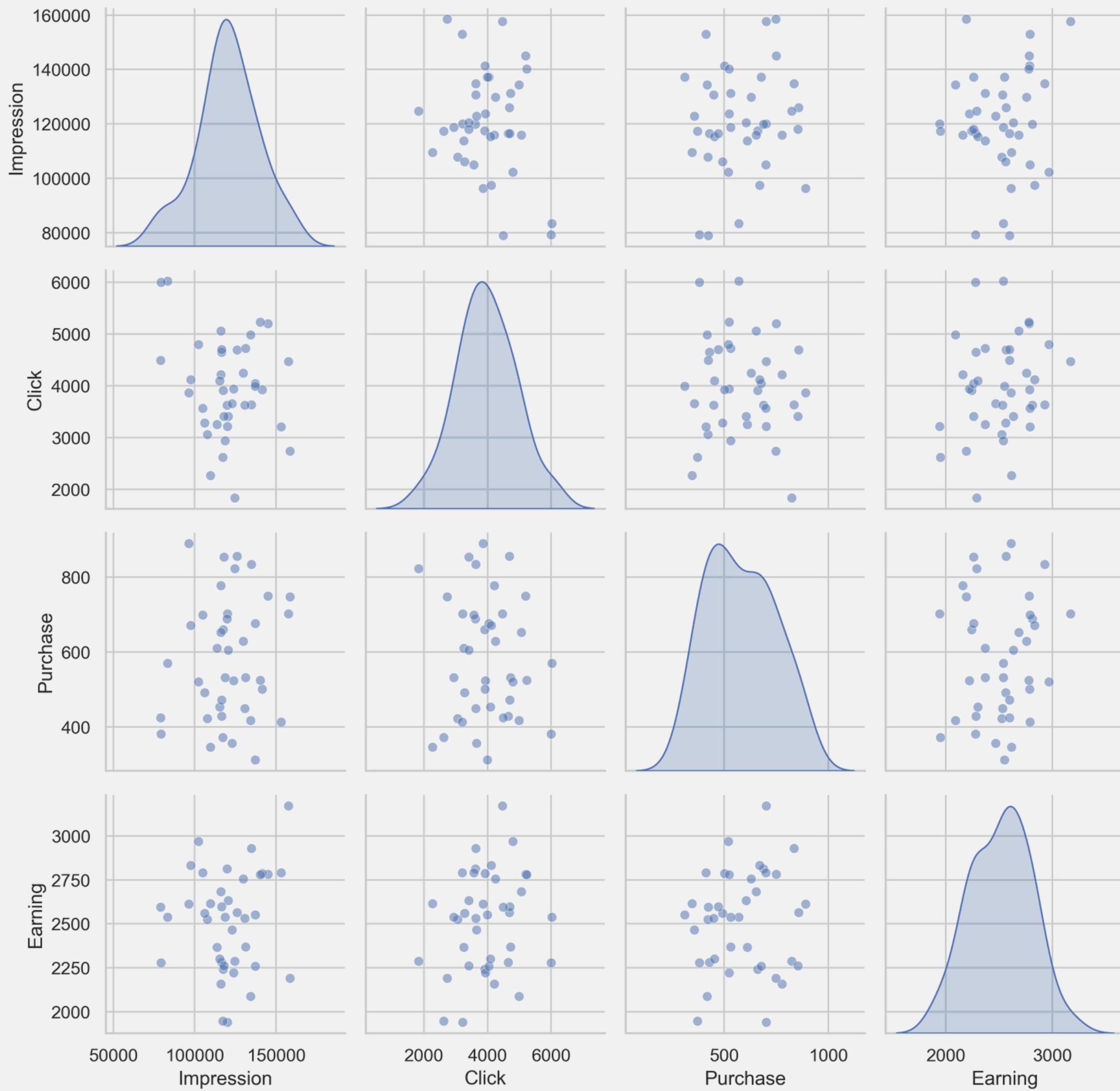
The pair plot shows that while impressions and clicks have weak or no linear correlations with earnings, a slight upward trend between purchases and earnings implies that purchases may be a more critical metric for driving earnings in the control group, with individual distributions peaking around 100,000 for impressions, 5,000 for clicks, 500 for purchases, and 1,750 for earnings.

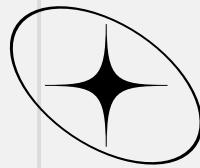




TEST GROUP DISTRIBUTIONS

The pair plot for the test group shows that while impressions peak around 140,000, clicks around 4,000, purchases around 500, and earnings around 2,750, there are **weak or no linear correlations between impressions and the other variables**, but a slight upward trend between purchases and earnings suggests **increasing purchases could boost earnings more effectively than increasing impressions or clicks**.

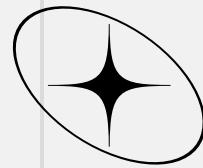




CONTROL GROUP CORRELATION MATRIX

The correlation matrix for the control group indicates that while **increasing impressions slightly boosts purchases (0.21)** and **earnings (0.11)**, clicks do not significantly translate into purchases (0.06) or earnings (-0.36), suggesting that **optimizing factors driving impressions could be more effective for improving ad performance.**

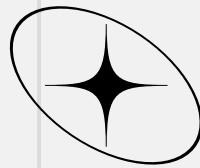




TEST GROUP CORRELATION MATRIX

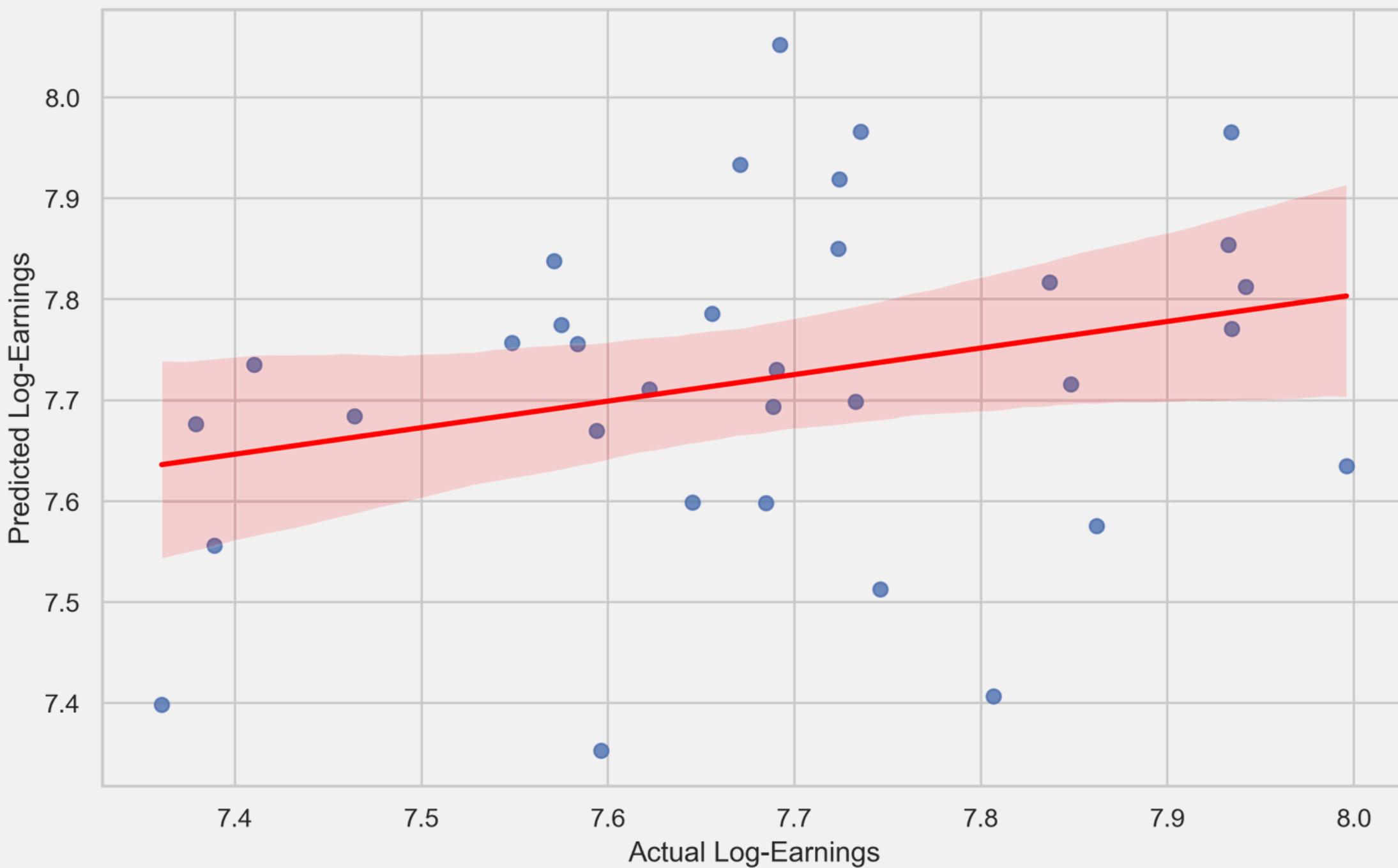
The test group's correlation matrix reveals weak positive correlations between impressions and purchases (0.14), impressions and earnings (0.11), and clicks and earnings (0.20), indicating that **impressions and clicks slightly boost earnings but have minimal impact on purchases.**

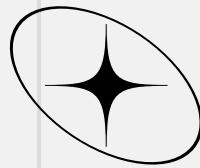




ACTUAL VS PREDICTED LOG-EARNING (LINEAR REGRESSION)

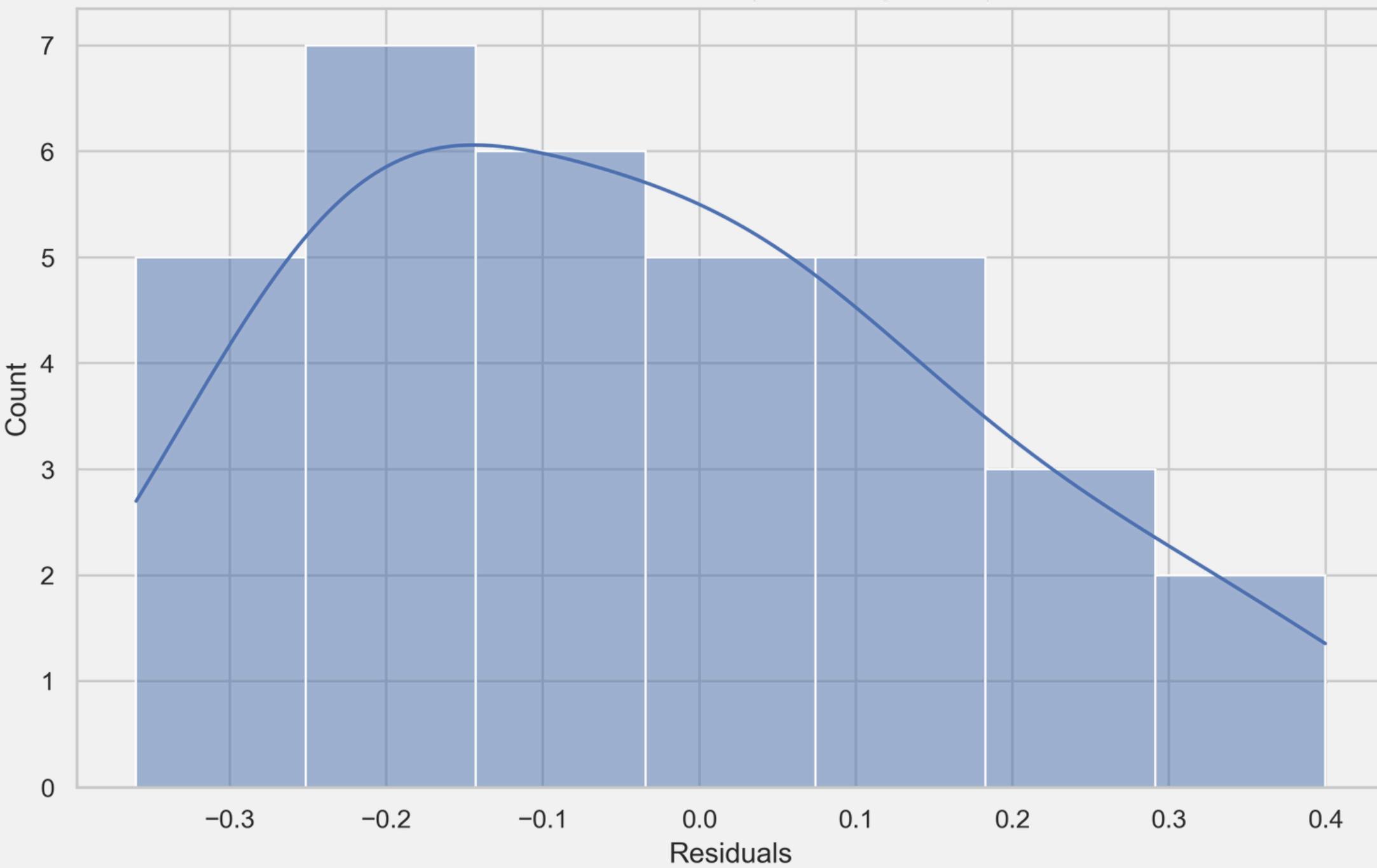
The scatter plot shows a positive correlation between actual and predicted log-earnings from a Linear Regression model. Still, significant residuals and a wide confidence interval indicate that the model only partially captures the variability in actual log-earnings.

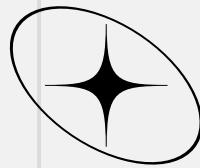




RESIDUALS DISTRIBUTION (LINEAR REGRESSION)

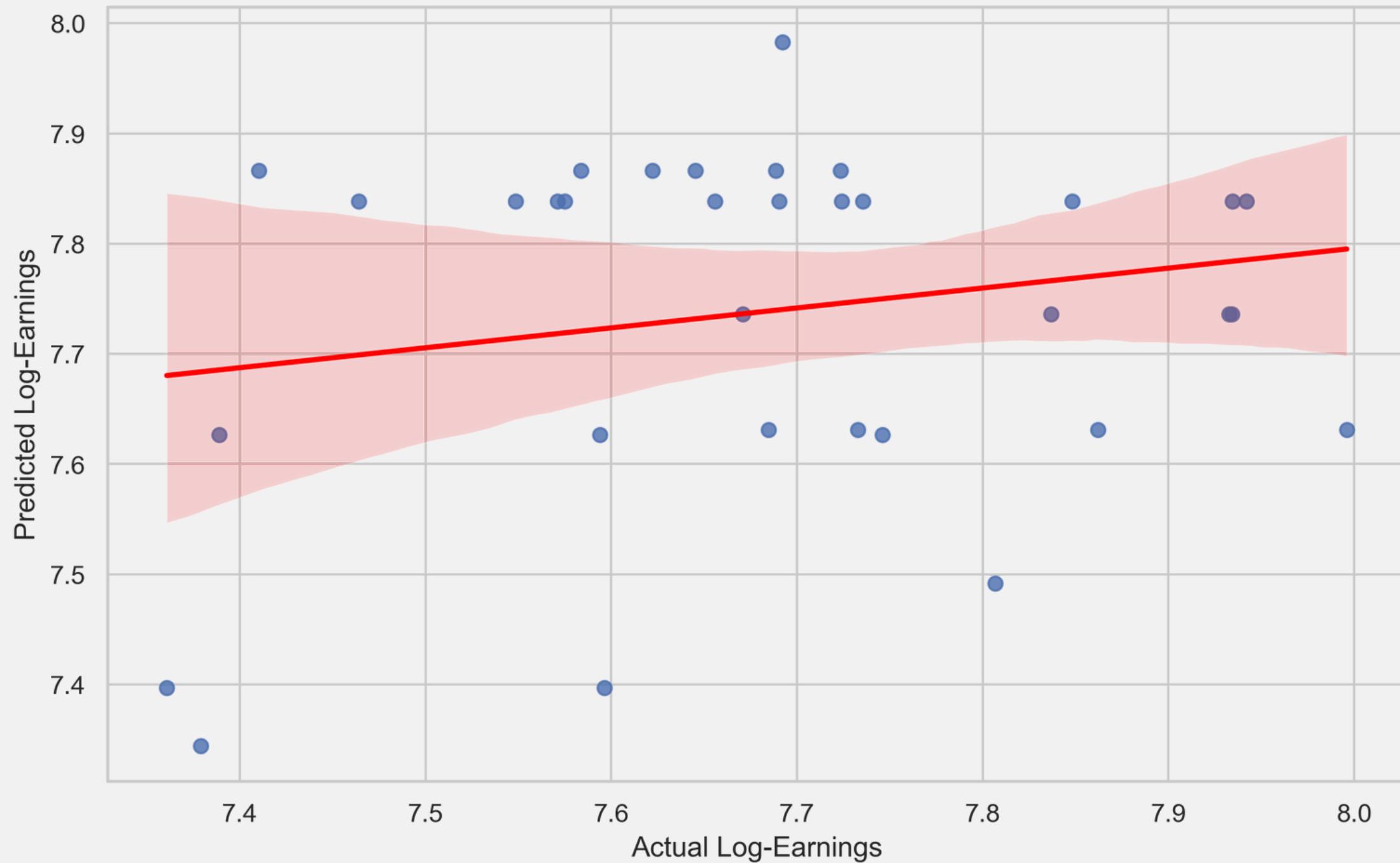
The residuals distribution indicates that while the Linear Regression model predicting log-earnings performs reasonably well, significant deviations and left skewness suggest it under-predicts in some cases and does not fully capture all influencing factors.

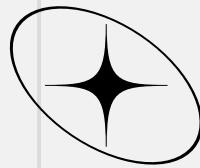




ACTUAL VS PREDICTED LOG-EARNING (DECISION TREE)

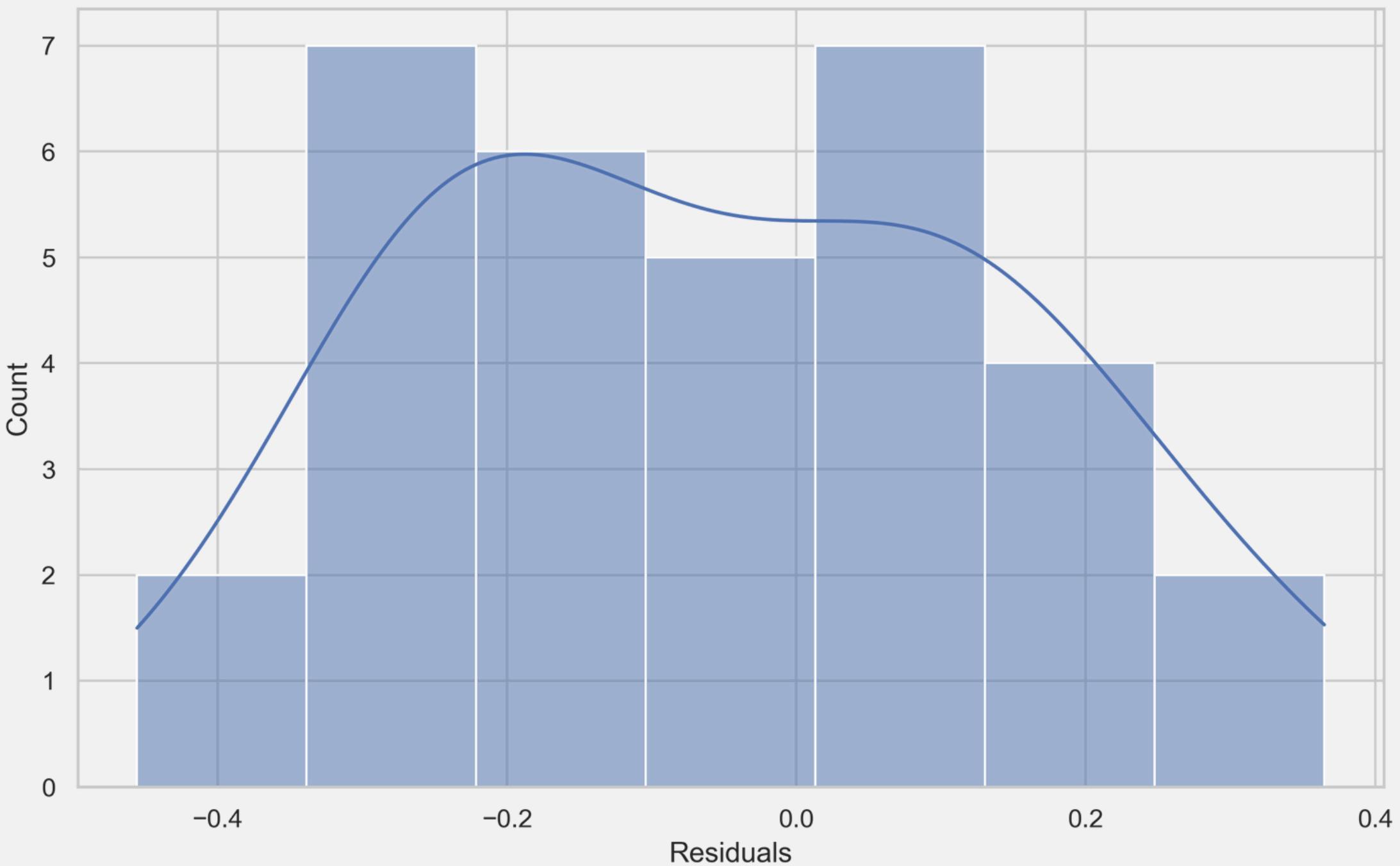
The scatter plot shows that the Decision Tree model's predictions for log-earnings align somewhat with actual values but exhibit significant variability and wider confidence intervals at the extremes, indicating moderate predictive power but less accuracy for high or low values.

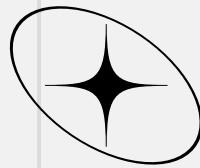




RESIDUALS DISTRIBUTION (DECISION TREE)

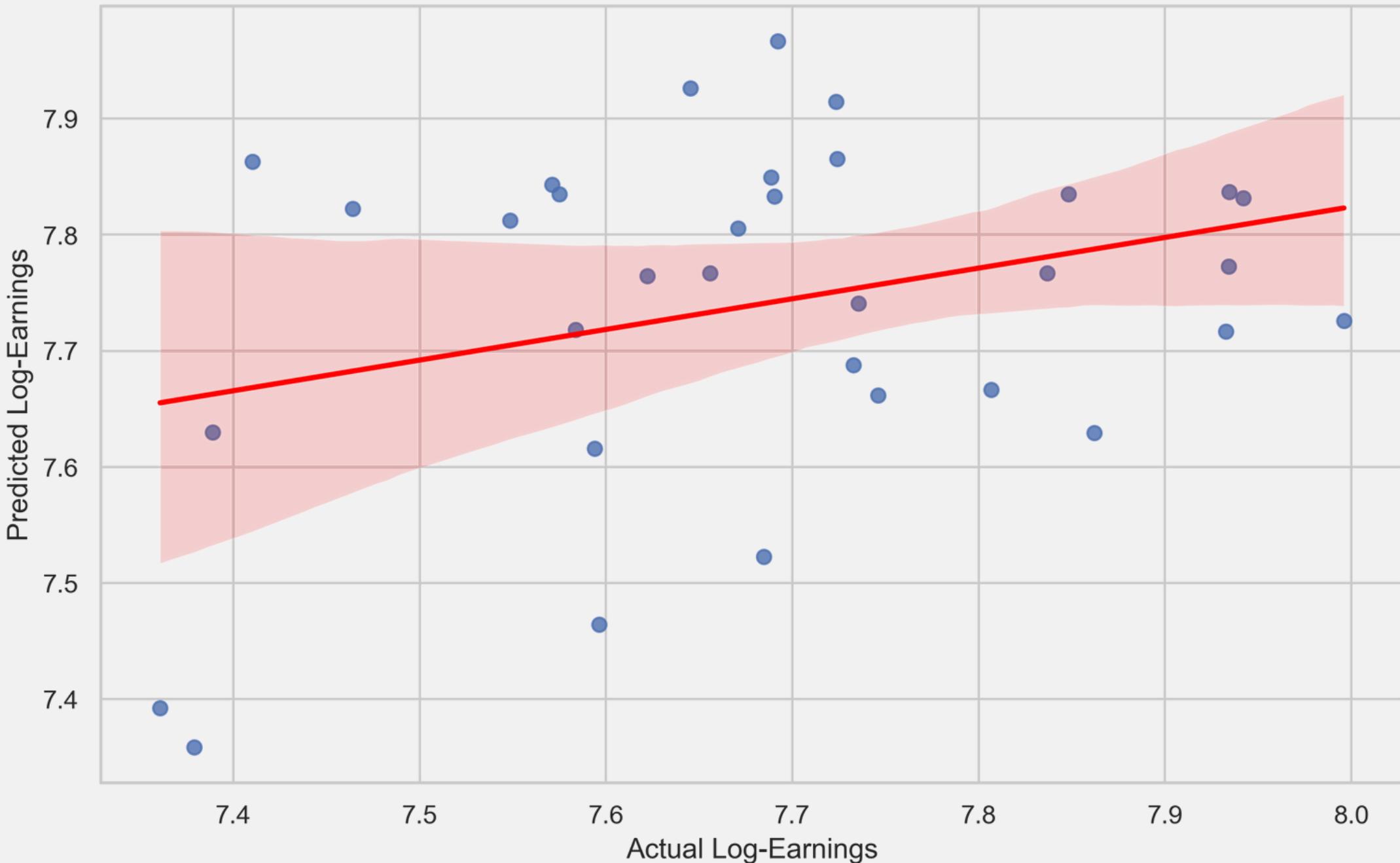
The histogram of residuals for the Decision Tree model predicting log-earnings shows a relatively even spread around zero with notable peaks at the extremes, indicating **balanced but inconsistent predictions** and highlighting the model's **significant prediction errors and limitations**.

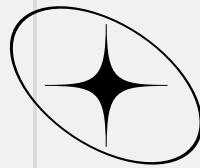




ACTUAL VS PREDICTED LOG-EARNING (GBT)

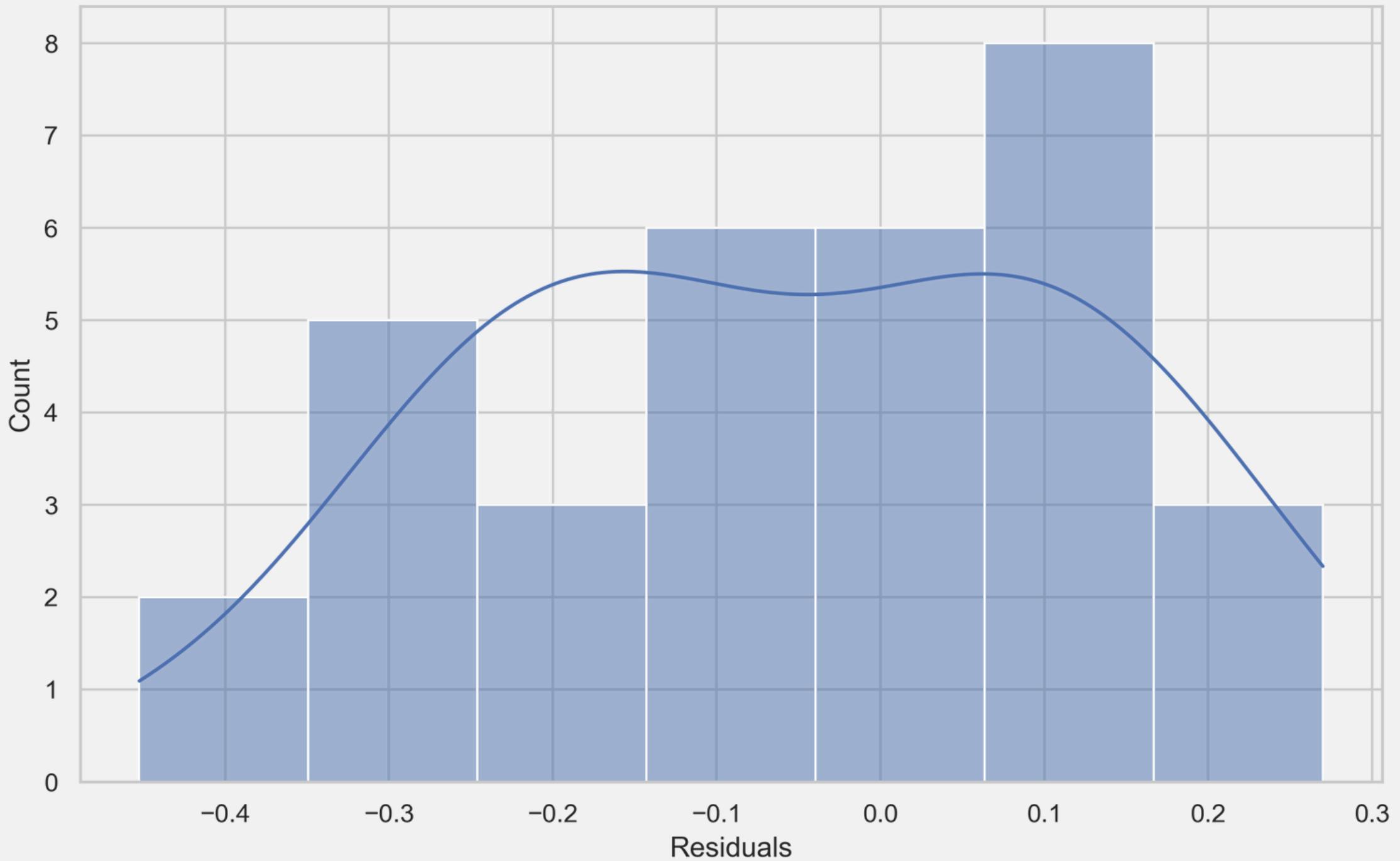
The scatter plot of actual versus predicted log-earnings with a Gradient Boosting Trees model shows a **growing prediction uncertainty for higher earnings**, as indicated by the widening red confidence interval, and highlights significant prediction errors, particularly for certain data ranges.

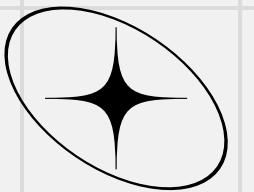




RESIDUALS DISTRIBUTION (GBT)

The histogram of residuals for the Gradient Boosting Trees model predicting log-earnings shows a balanced spread around zero but with significant errors at the extremes, highlighting the model's occasional substantial prediction inaccuracies and its limitations in capturing all data nuances.

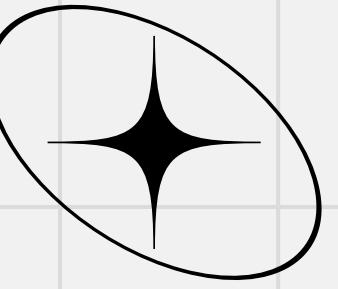




CONCLUSION

The analysis of predictive models—Linear Regression, Decision Tree Regression, and Gradient Boosted Trees (GBT) Regressor—reveals that the GBT model is the most effective for estimating log-earnings due to its ability to model complex, non-linear relationships. While impressions and clicks show weak correlations with earnings, suggesting their limited impact on financial outcomes, the GBT model offers a more accurate prediction by effectively handling variability and capturing underlying trends. Future work should focus on incorporating additional variables and optimizing the GBT model through cross-validation and hyperparameter tuning to enhance forecasting accuracy further and provide deeper insights into the factors driving earnings.





THANK YOU

