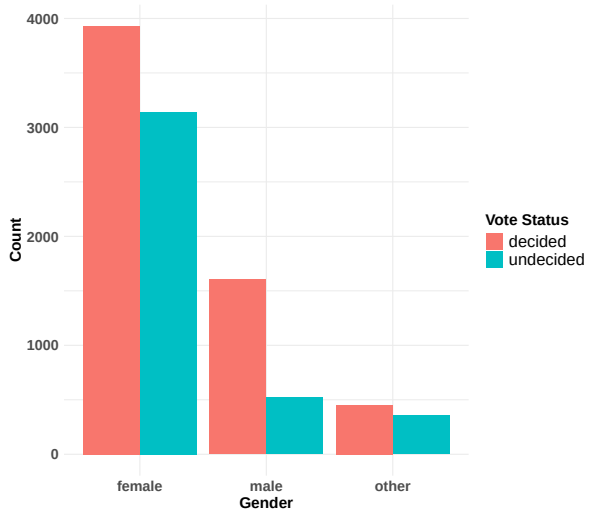
Analyzing Voter Indecision in an Election Campaign

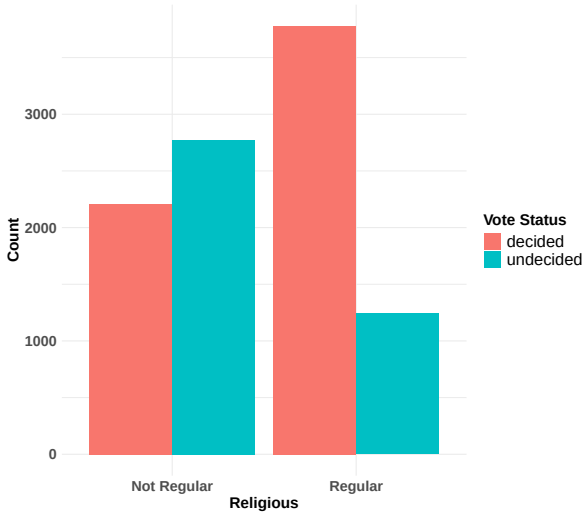
Saakshi Agarwal

This report examines voter indecision using a dataset of approximately 10,000 observations containing voter attributes related to demographics, engagement, and political behavior. The key variables influencing indecision are Gender, Education, and Religious Affiliation, as gender dynamics impact decision-making, education correlates with political engagement, and religious affiliation strongly predicts voting behavior.

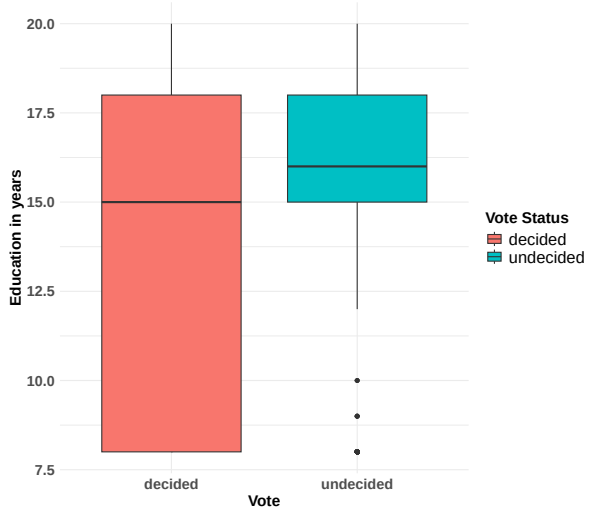
Figure 1a show that among 4014 undecided voters, with 3140 being female, indicating potential differences in political engagement across genders. Figure 1b shows that among 4980 non-regular religious 2772 of them are undecided, suggesting that religious affiliation strengthens political convictions. Figure 1c and Table 1 confirms that undecided voters tend to have slightly higher education levels than decided voters. The median education level for undecided voters is 16 years, compared to 15 years for decided voters, with a slightly higher mean. This trend suggests that higher-educated individuals may take longer to evaluate their choices before deciding.



(a) Gender by Vote status



(b) Religious by Vote status



(c) Distribution of Education by Vote Status

Figure 1: Bivariate analysis for Vote againts Key Variables

Table 1: Descriptive analysis for Education (years) based on Vote

Vote	Mean (years)	Median (years)	Q1 (years)	Q3 (years)	Min (years)	Max (years)
Decided	14.05	15	8	18	8	20
Undecided	16.62	16	8	18	8	20

To predict voter indecision, the dataset was split into 65% training and 35% testing for unbiased model evaluation. Several classification models were tested, and LightGBM was selected for its superior performance, achieving 91.5% recall (sensitivity) and 97.7% positive predictive value (PPV) [Figure 2]. High recall ensures that most undecided voters are identified, making outreach efforts more effective, while high PPV ensures that flagged undecided voters are truly at risk of disengagement. The ROC AUC score of 96.2% confirms the model's ability to differentiate between undecided and decided voters.

Figure 2: Model comparisom for summary of classification metric

Table 2: Confusion Matrix for LightBGM Model (Based on Test Data)

	Actual: No	Actual: Yes
Prediction: No	1952	44
Prediction: Yes	144	1361

The confusion matrix [Table 2] shows the model correctly classified 1,952 decided voters and 1,361 undecided voters, misclassifying 144 decided voters as undecided and 44 undecided voters as decided. The overall accuracy was 94.6%, and the test ROC AUC score of 96.2% [Figure 3] further validates the model's effective differentiation between undecided and decided voters. As shown in Table 3, the model achieved a recall of 93.1%, successfully identifying most undecided voters. A precision (PPV) of 97.8% suggests that most voters flagged as undecided were correctly classified. While specificity was slightly lower (96.9%), the trade-off is acceptable, as reducing false negatives (missed undecided voters) is a priority. The negative predictive value (NPV) of 90.4% suggests that some undecided voters were misclassified, but the model provides a well-balanced approach.

Table 3: Summary of Classification Metric (Based on Test Data)

Metric	Estimate
Accuracy	0.946
Sensitivity	0.931
Specificity	0.969
Positive Predictive Value (ppv)	0.978
Negative Predictive Value (npv)	0.904

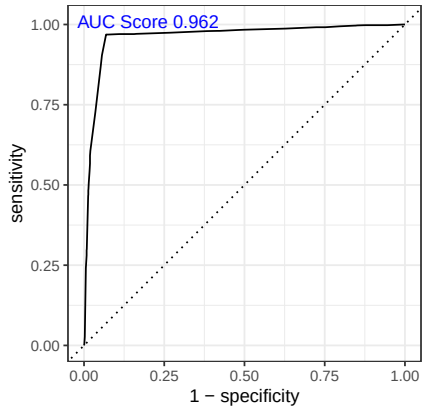


Figure 3: ROC Curve for Model Accuracy (Test Data)

Based on these insights, the campaign should prioritize outreach to females, non-religious individuals, and higher-educated voters, who are more likely to be undecided. Tailored messaging and engagement strategies for these groups can increase voter persuasion. LightGBM's strong recall and precision make it a valuable tool for identifying undecided voters, enabling campaign teams to optimize engagement initiatives and enhance voter turnout.