LEAD SCORING CASE STUDY

Presented by:
Aravind Peddinti
Arnave Pradeep
Anuja Patil

CASE STUDY DESCRIPTION

- X Education, an online course provider for students, industry professionals etc., attracts visitors through marketing on websites and search engines like Google. Professionals landing on their website can browse courses, watch videos, or fill out a form with their contact information to become leads. The company also receives leads via referrals. Once leads are acquired, the sales team follows up through calls and emails to convert them into paying customers.
- Currently, X Education's lead conversion rate is low, with only about 30% of leads converting. For instance, out of 100 leads acquired daily, only 30 typically convert. This inefficiency impacts the company’s ability to optimize its resources and boost revenue.
- To improve efficiency and increase conversions, X Education aims to identify 'Hot Leads,' or leads with the highest potential for conversion. By focusing on these leads, the sales team can prioritize their efforts, reducing unnecessary outreach and potentially increasing the conversion rate.

AGENDA

X Education has tasked us with building a model to identify the most promising leads, assigning a lead score to each based on their likelihood of conversion. Our goal is to ensure that leads with higher scores have a higher chance of converting into paying customers, while those with lower scores are less likely to convert. The CEO has set a target to significantly improve the current conversion rate, aiming for around 80%, and our focus will be on developing a robust and accurate scoring system.

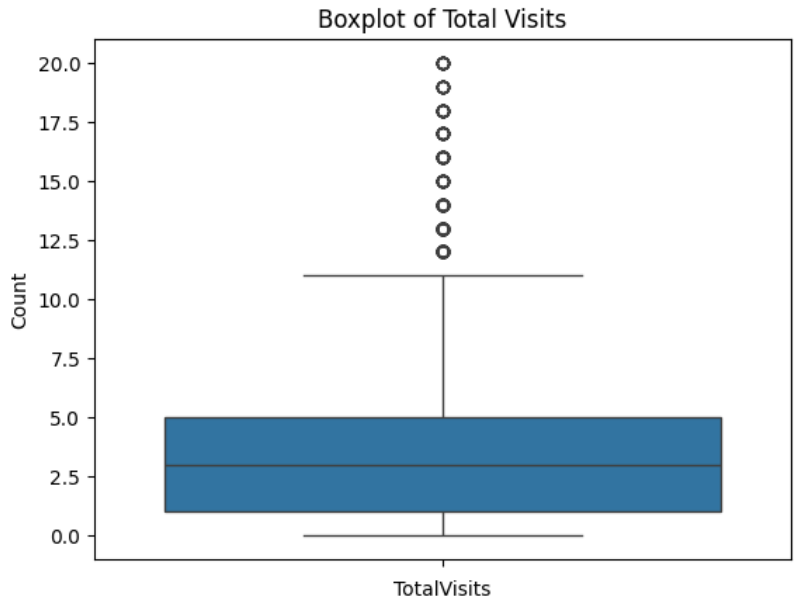
APPROACH

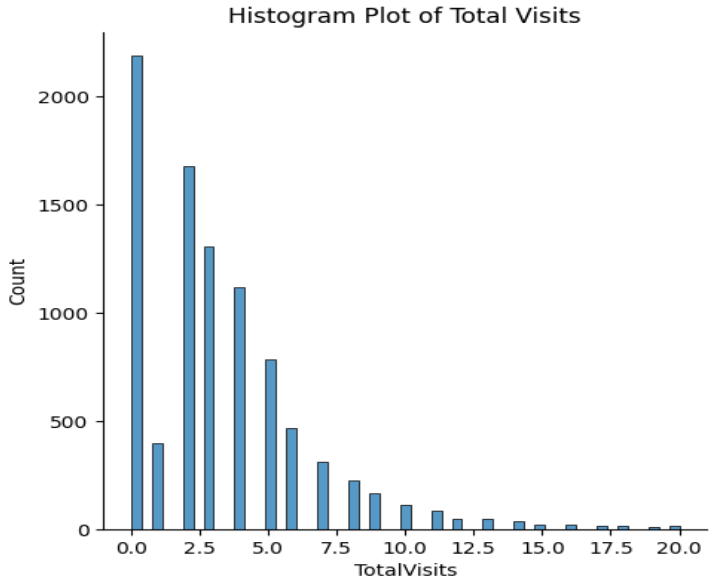
- Sourcing the data
- Reading and understanding the data
- Data Cleaning
- EDA
- Scaling the data
- Splitting the data into test and training set
- Model building
- Model evaluation

IMPORTING, UNDERSTANDING AND CLEANING THE DATA

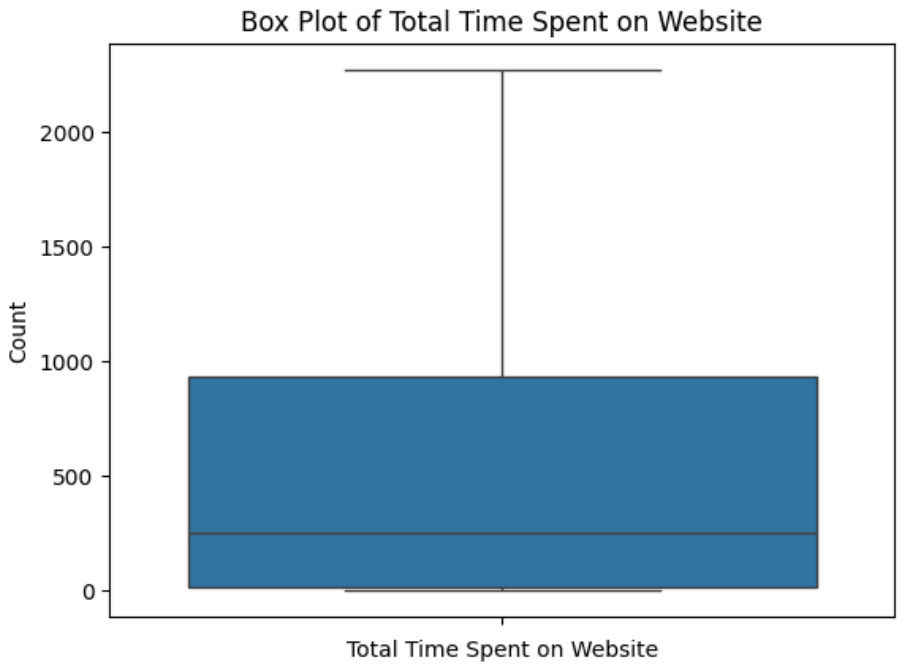
- Import the data from CSV data source.
- Removing columns with missing data (missing >30% of data)
- Handling Outliers
- Impute missing values
- Removing redundant columns
- Exploratory Data Analysis – Univariate, Bivariate and Multi-variate analysis

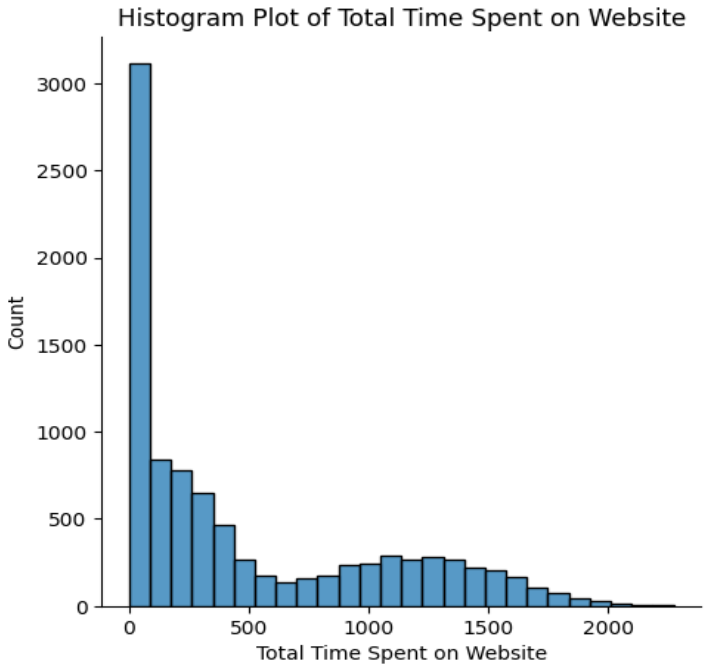
Exploratory Data Analysis



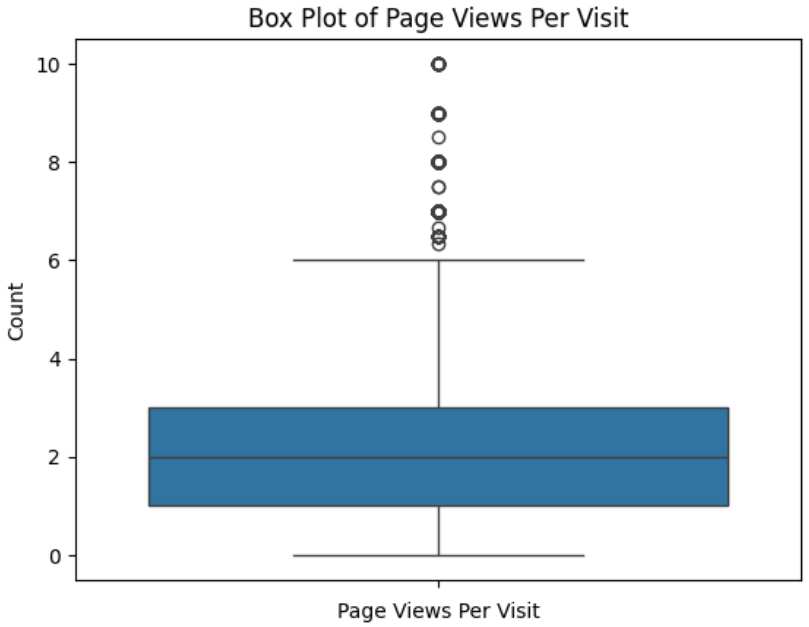


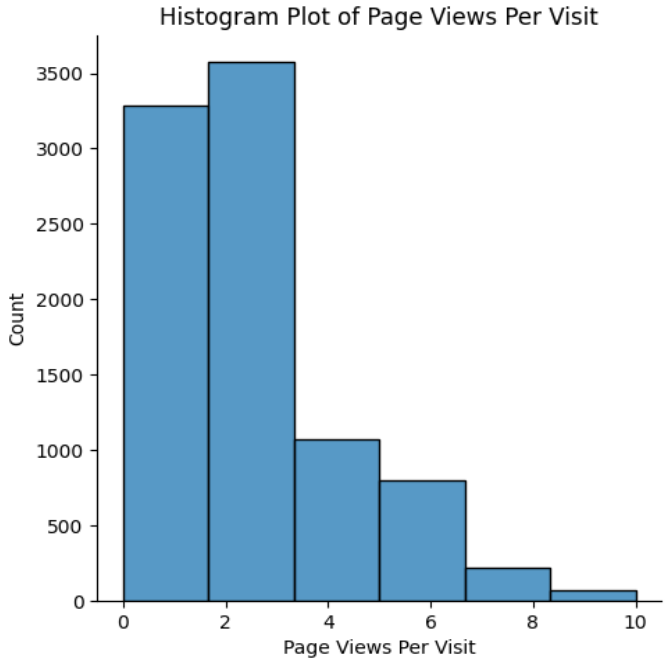
Exploratory Data Analysis



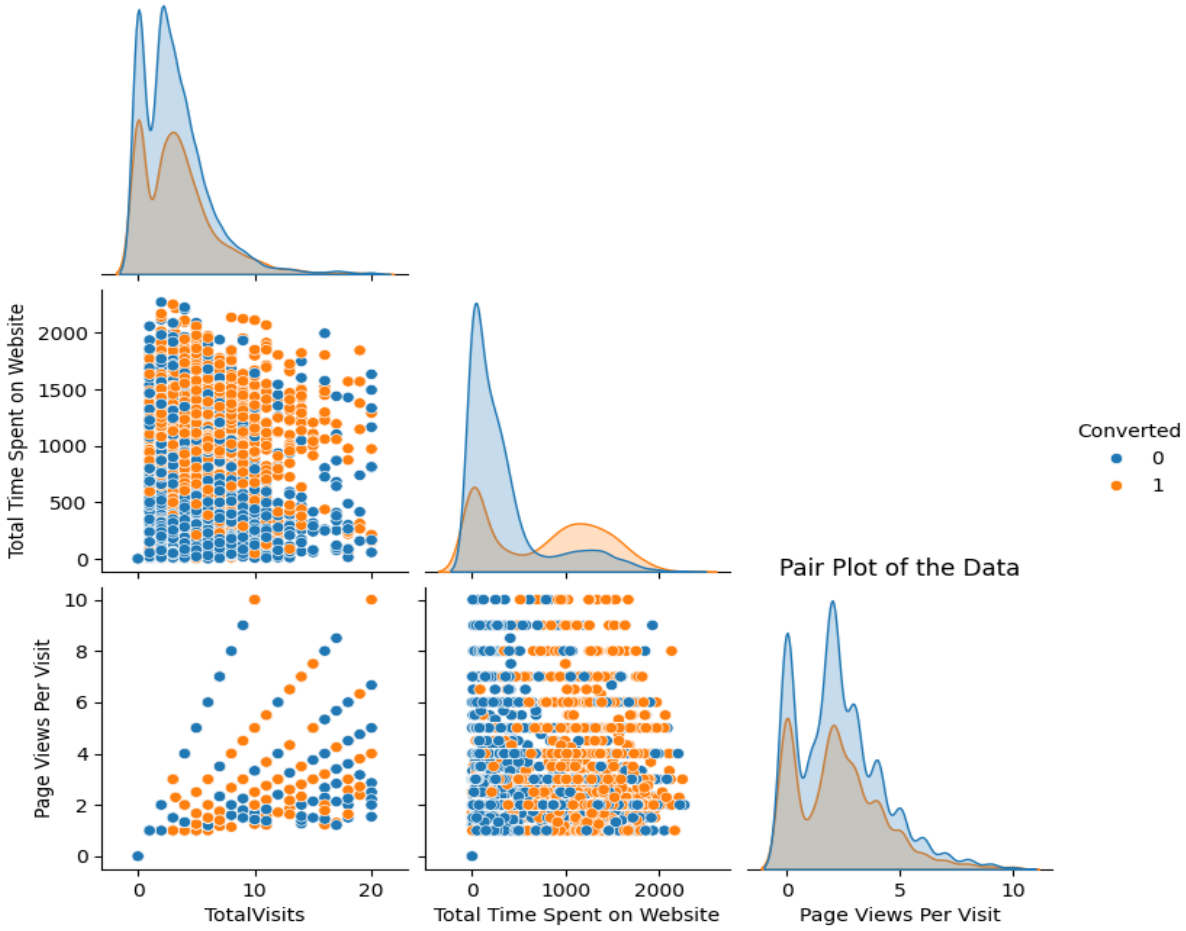


Exploratory Data Analysis

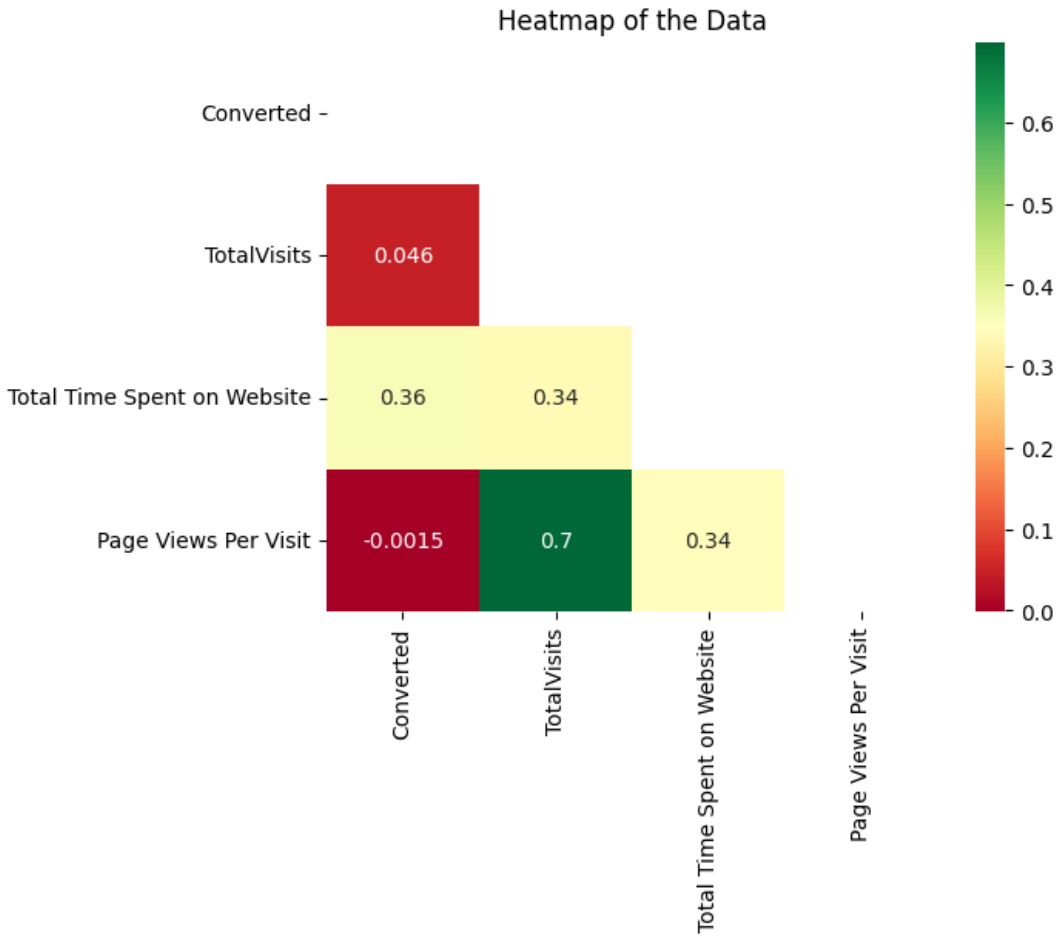




Exploratory Data Analysis



Exploratory Data Analysis



MODEL BUILDING

- Create dummy variables
- Feature scaling
- Train-Test split
- Dimensionality reduction using RFE
- Building model using Logistic Regression (using statsmodels.api.Logit())
- Feature elimination based on p-values and VIF

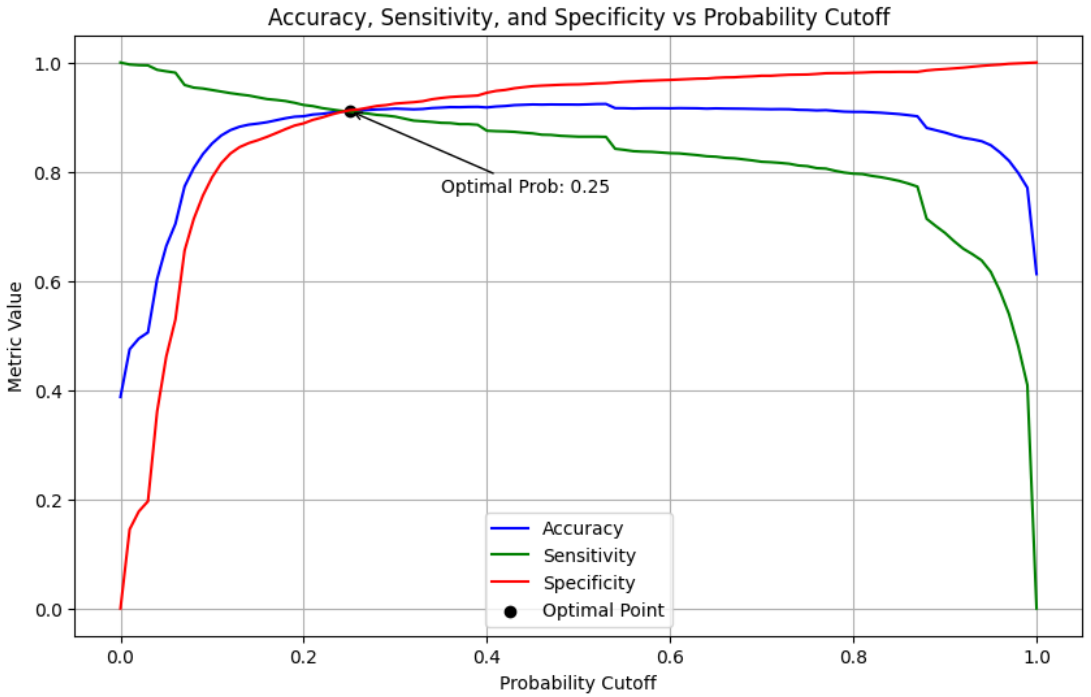
MODEL EVALUATION

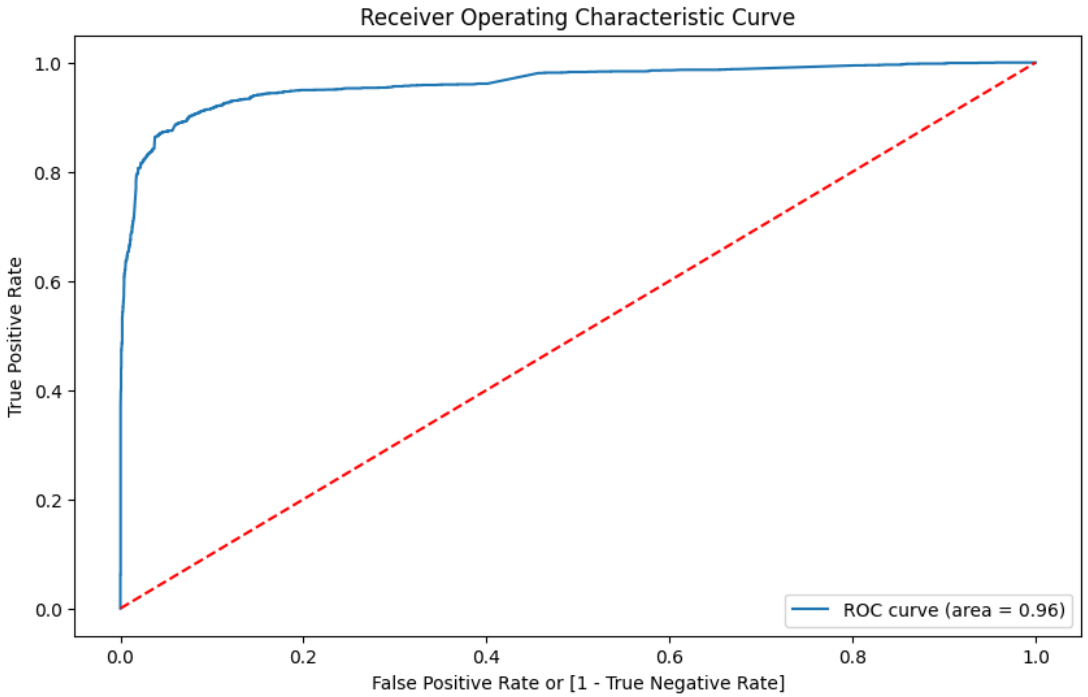
- Calculated Accuracy, Precision, Recall, F1 Score, Sensitivity and Specificity.
- Plotted Accuracy, Sensitivity and Specificity Curve and Receiver Operating Characteristics (ROC) curve
- Calculated optimal probability threshold
- Plotted the confusion matrix

IMPORTANT FEATURES OF THE DATASET

- Total Time Spent on Website
- Lead Source_Welingak Website
- Occupation_Missing
- Tags_Busy
- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Tags_Ringing
- Tags_Will revert after reading the email
- Tags_switched off
- Last Notable Activity_Olark Chat Conversation
- Last Notable Activity_SMS Sent

Accuracy, Sensitivity and Specificity Curve and (ROC) Curve

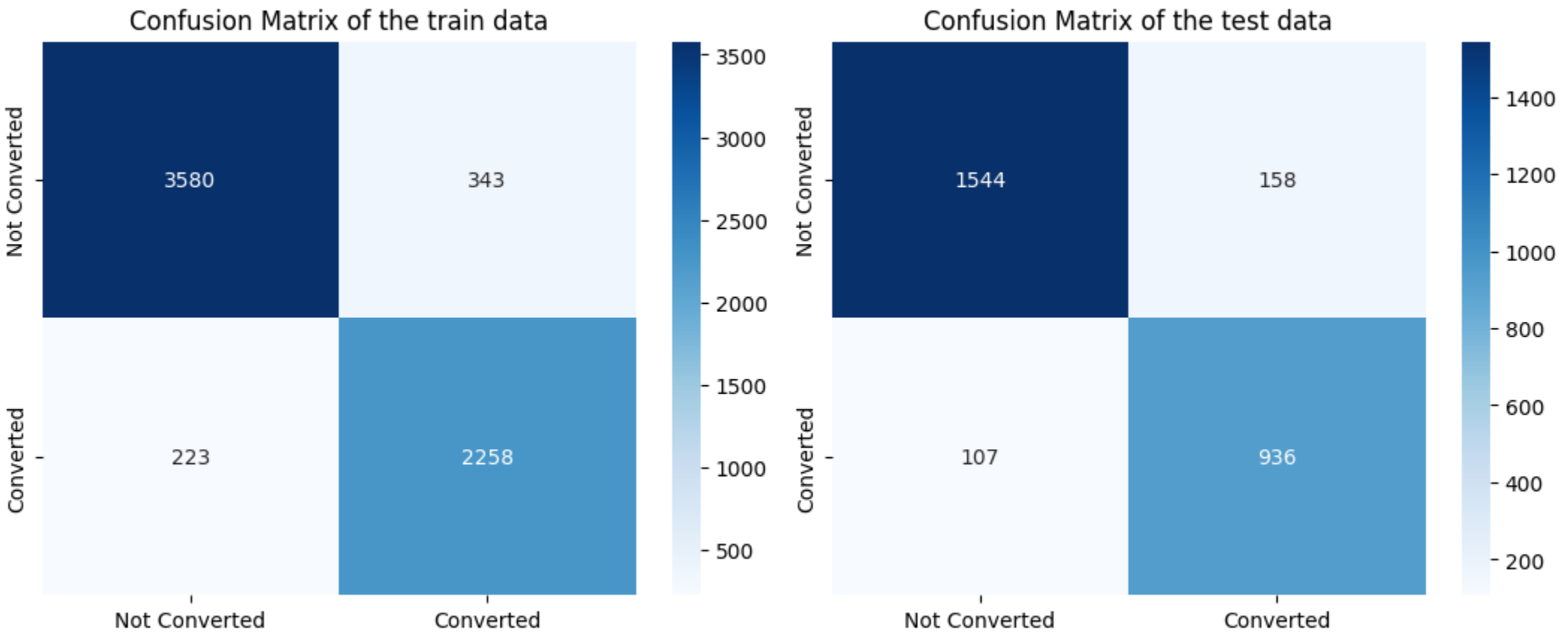




Evaluation Parameters

Metric/Data	Training Data	Test Data
Accuracy	0.91	0.9
Precision	0.87	0.86
Recall	0.91	0.9
F1 Score	0.89	0.88
Sensitivity	0.91	0.91
Specificity	0.91	0.9
ROC AUC	0.96	

CONFUSION MATRIX



CONCLUSION

- This model performs well on both the training and test data. The accuracy, precision, recall, and F1 scores are all high, indicating a good model.
- The sensitivity and specificity for both the training and test data are also high, indicating that the model is able to correctly classify the converted and not converted leads.
- The ROC curve and AUC score are also high, indicating that the model is very efficient. Overall, this model is good for predicting lead conversion.

SUMMARY

As seen from model's important features, it is important to assign a tag to each of the leads. The leads generated by opening Welingak website and redirected to X Education are most likely to get converted.

Monitor people who spent more time on X Education website and contact them as they are likely to get converted. We can increase the ads on the Welingak website to increase the no. of redirections and also to increase the view time of the website.

There is also need to strategize whom to call for lead conversion, such as avoiding repetitive calls to people whose phones were switched off or keeps on ringing during a call and avoiding contacting people that haven’t mentioned their occupation. Also monitor people who are not joining competitors as they would most likely to get converted.