

# LEAD SCORING CASE STUDY

---

SHOAIB  
MOHAMMED

SUNITHA

SHUBHAM  
THAKUR

# Problem Statement

- ❖ **Company Overview:** X Education sells online courses to industry professionals.
- ❖ **Lead Generation:** Professionals land on the website, browse courses, fill forms, or watch videos. Leads are also acquired through referrals.
- ❖ **Current Process:** Sales team follows up with leads via calls and emails, with a conversion rate of around 30%.
- ❖ **Problem:** Despite generating many leads, the conversion rate is low (30%), with only about 30 out of 100 leads converting.
- ❖ **Objective:** Identify "Hot Leads" to improve conversion rates by focusing on the most promising leads.
- ❖ **Goal:** Build a model to assign a lead score, targeting an 80% conversion rate, where higher scores indicate higher conversion potential.

# Objective

---

- **Objective:** Build a model to assign a lead score between 10-100, helping identify "Hot Leads" to improve the conversion rate.
- **Target:** The CEO aims for an 80% lead conversion rate.
- **Future-Proofing:** The model should be robust enough to handle future constraints such as:
  - Managing peak times effectively.
  - Optimizing the use of full manpower.
  - Developing strategies for post-target scenarios.
- **Outcome:** The lead scores will guide the sales team to focus on high-potential leads, ensuring efficient resource allocation and higher conversion rates.



# APPROACH

## 1. Introduction to Lead Scoring

## 2. Data Understanding and Preparation

- I. Overview of the Dataset
- II. Data Cleaning and Handling Missing Values
- III. Outlier Detection and Treatment
- IV. Feature Engineering
- V. Dummy Variable Creation for Categorical Data

## 3. Exploratory Data Analysis (EDA)

- I. Univariate and Bivariate Analysis
- II. Understanding Key Predictors
- III. Correlation Analysis

## 4. Model Building

- I. Logistic Regression Model Development
- II. Recursive Feature Elimination (RFE) for Feature Selection

## 5. Model Evaluation

- I. Confusion Matrix and Performance Metrics
- II. ROC Curve and AUC Score Analysis
- III. Optimal Cutoff Selection

## 6. VIF Analysis

- I. Multicollinearity Check

- II. Addressing Multicollinearity Issues

## 7. Final Model Interpretation

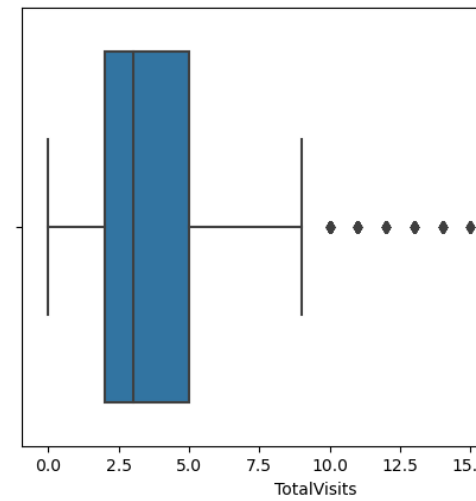
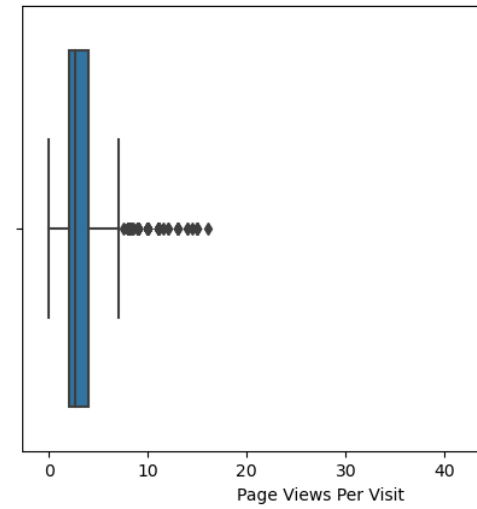
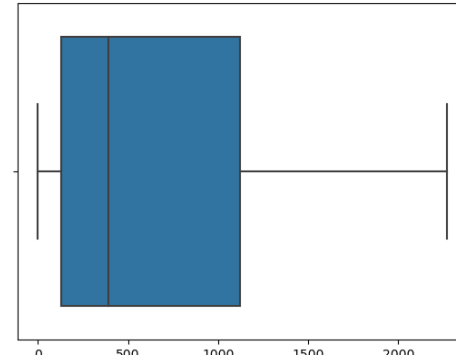
- I. Significant Predictors and Coefficients
- II. Final Model Performance Metrics

## 8. Conclusion and Recommendations

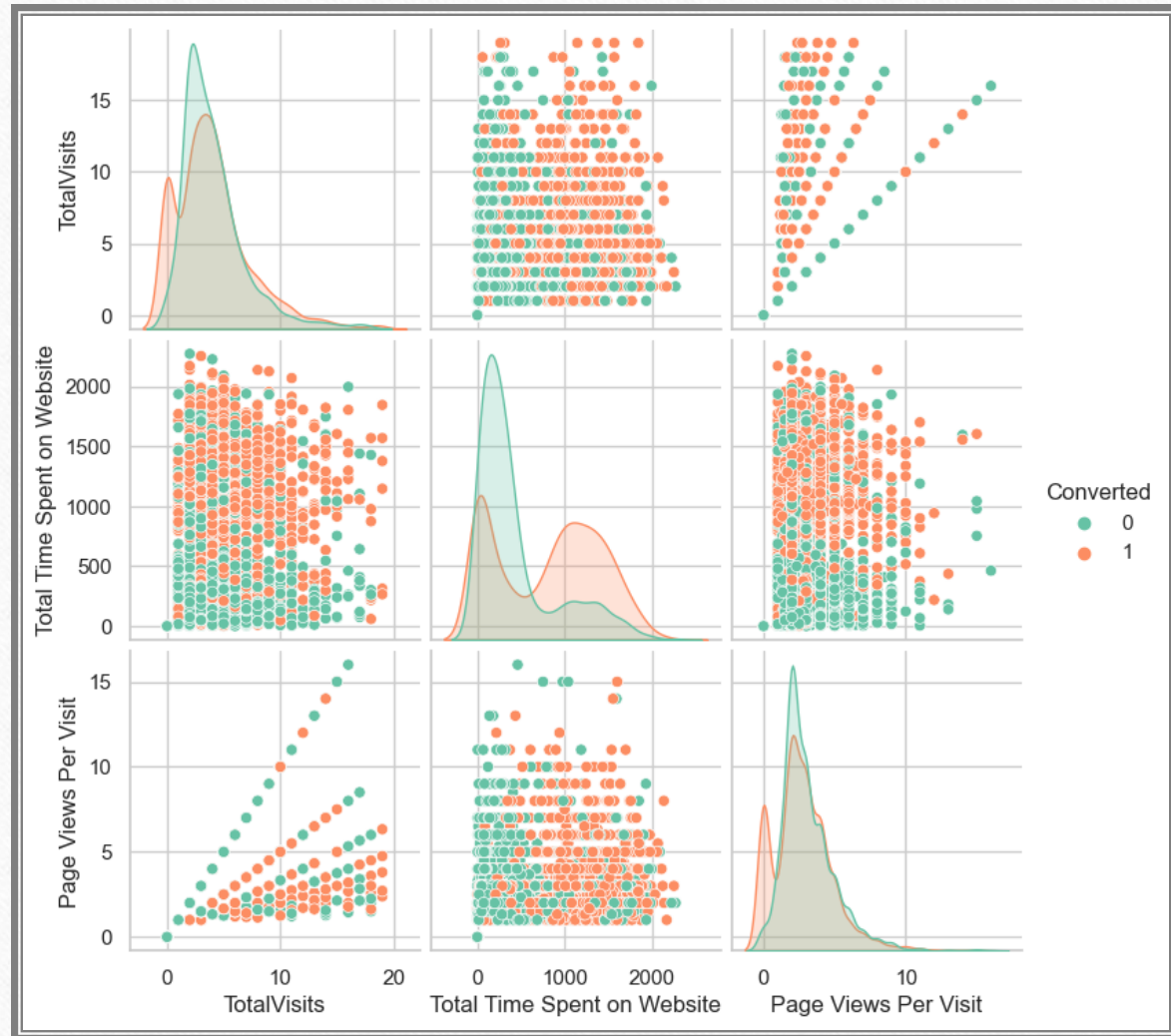
- I. Summary of Findings
- II. Recommendations for Implementation
- III. Future Considerations and Constraints

# OUTLINERS

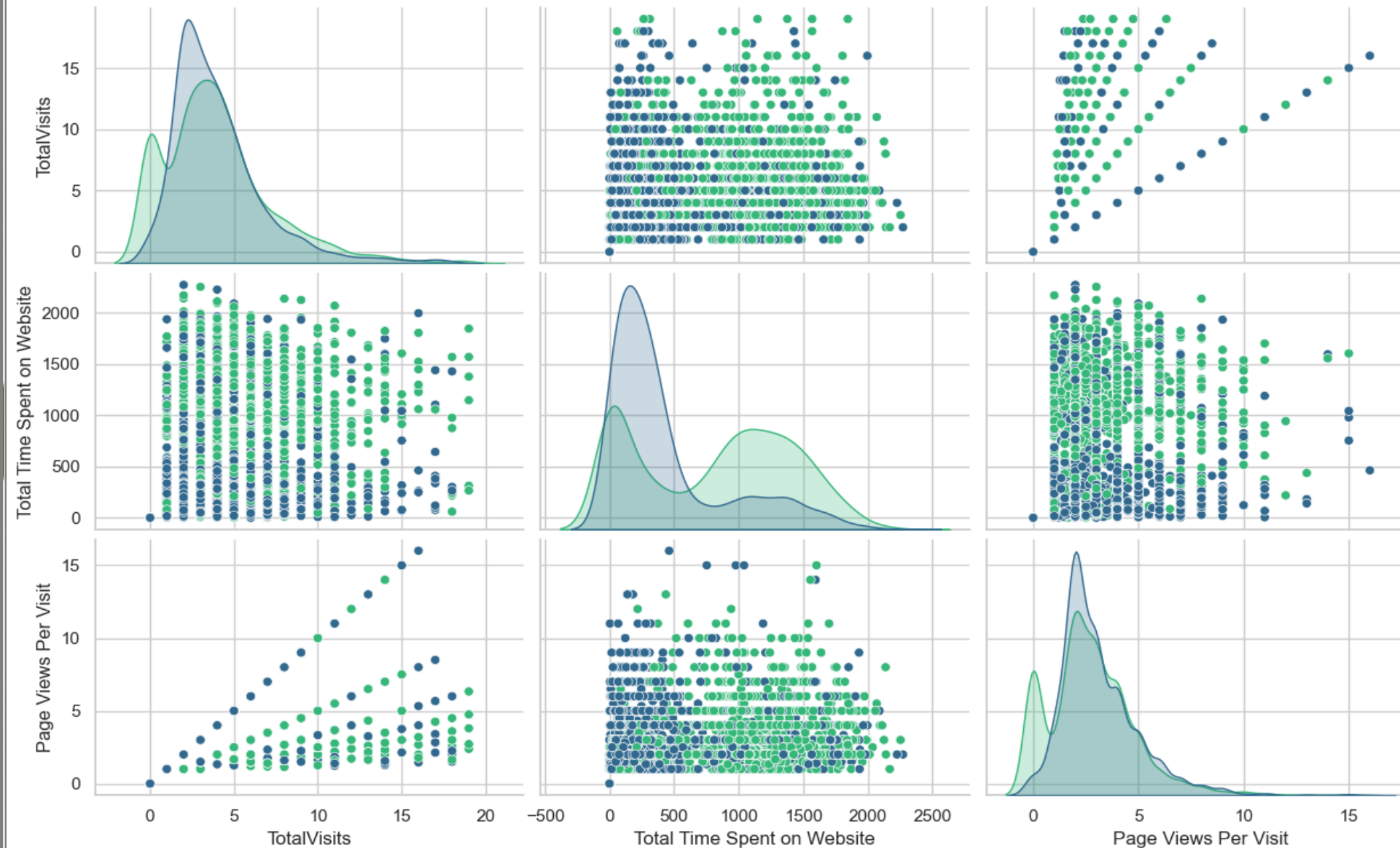
---



# PLOTS

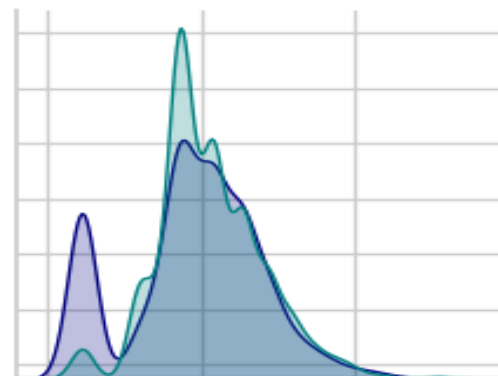
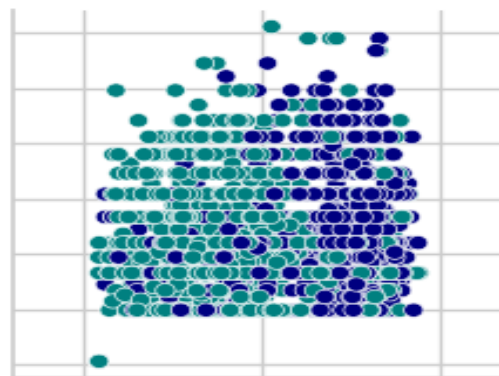
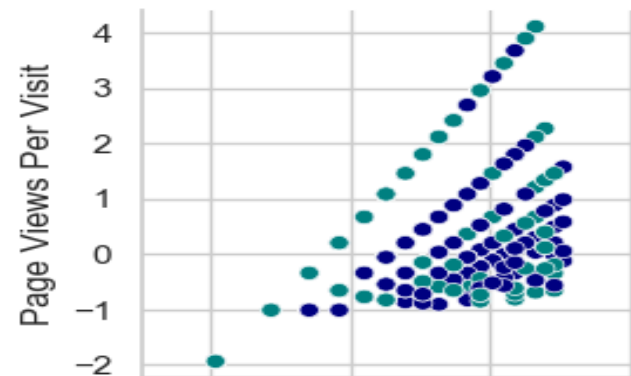
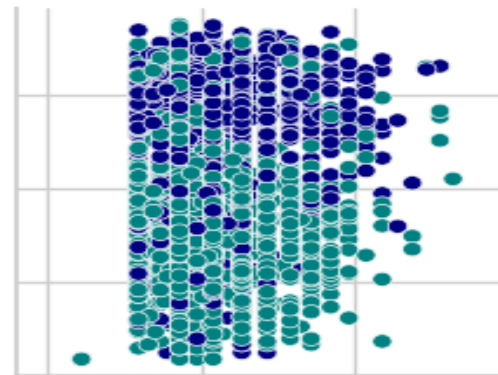
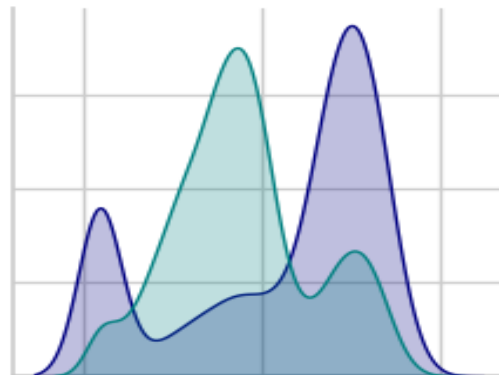
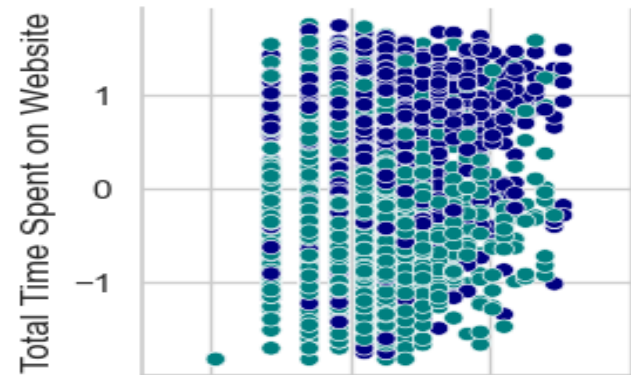
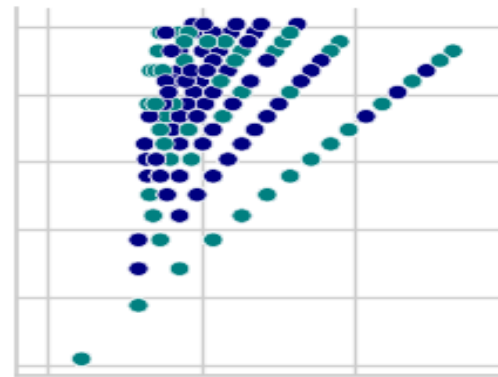
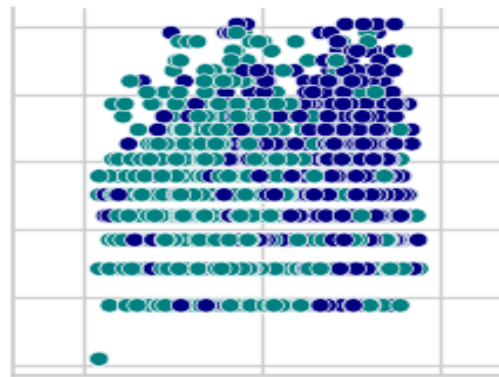
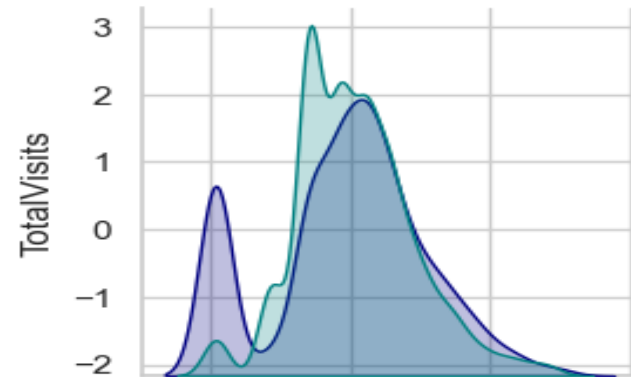






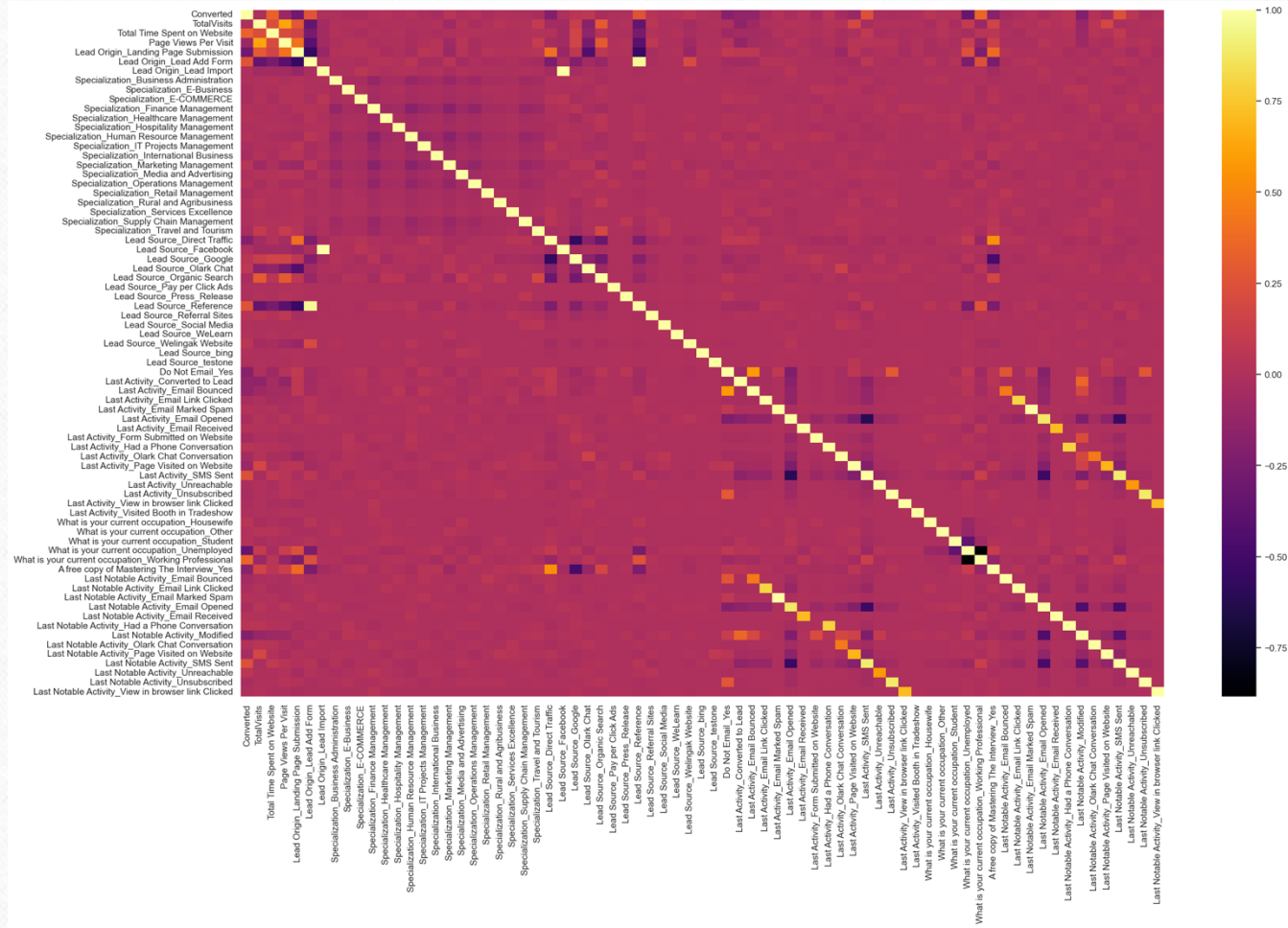
Converted

- 0
- 1



Converted  
 ● -1.0216046540388193  
 ● 0.9788522360842744

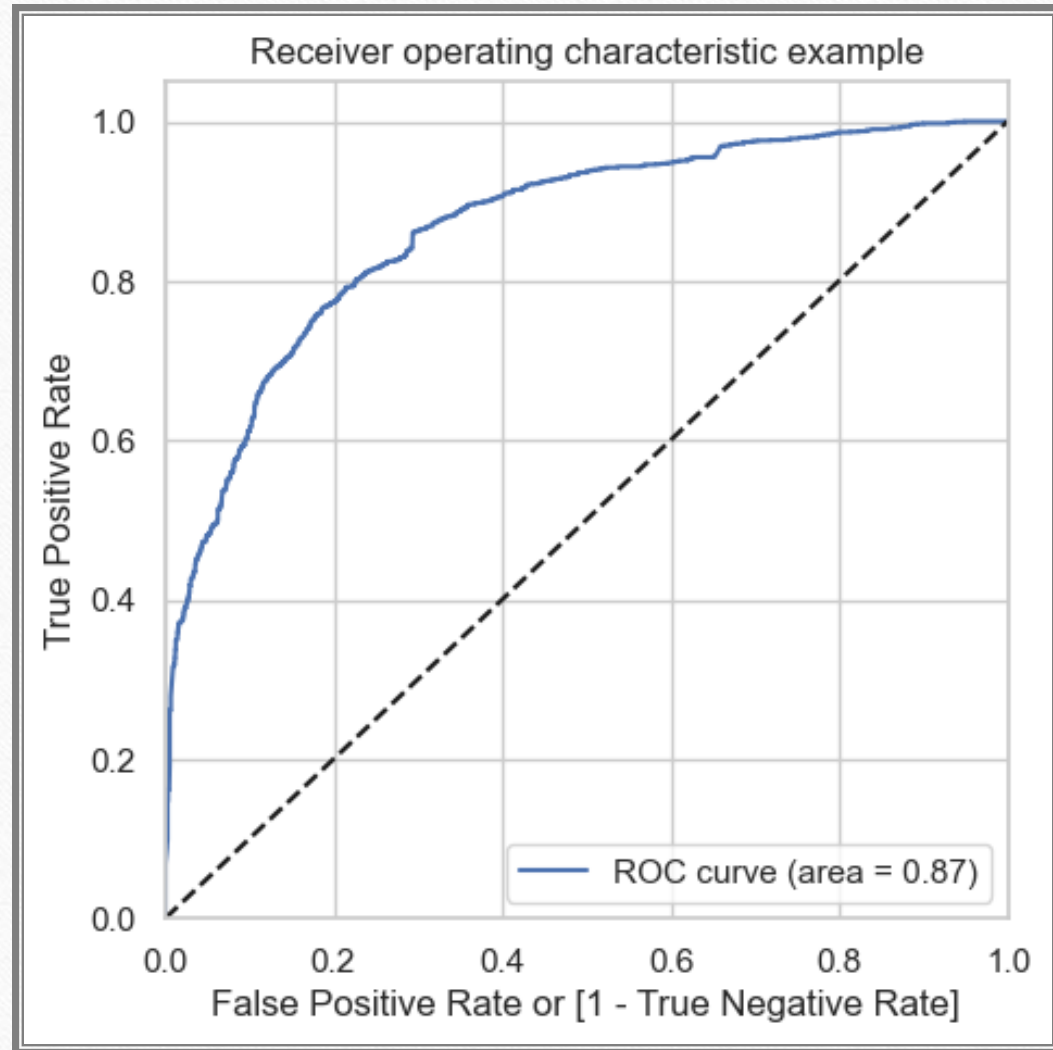




# ROC CURVE

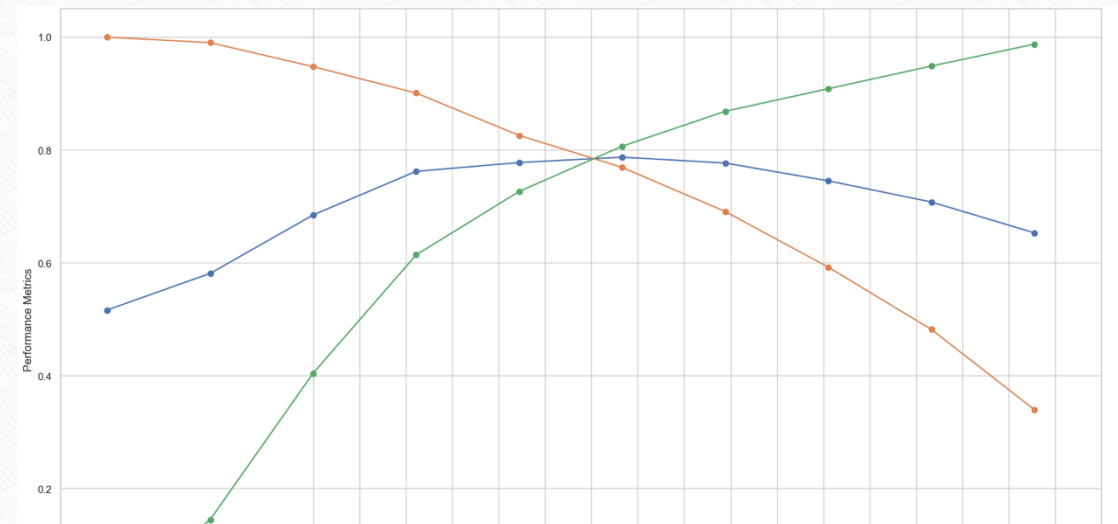
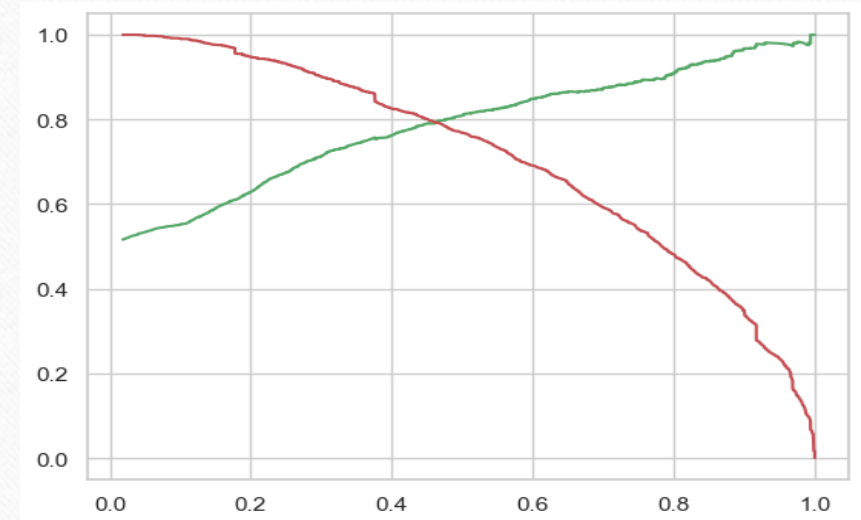


The area under the ROC curve is 0.87.



# EVALUATION PLOTS

- A threshold of 0.47 represents the optimal balance between precision and recall. Therefore, any prospect lead with a conversion probability higher than 47% can be confidently considered a hot lead.





# MODEL EVALUATION

TRAIN SET	TEST SET
•Accuracy: 78.7%	Accuracy: 79.3%
•Sensitivity: 76.9%	Sensitivity: 76.9%
•Specificity: 80.7%	Specificity: 81.7%
•Precision: 79.8%	Precision: 80.6%
•Recall: 78.9%	Recall: 76.9%

# Key Insights

---

- **Important Predictors:** Features like "TotalVisits," "Total Time Spent on Website," and certain categorical variables like "Lead Origin" and "Specialization" were identified as significant predictors of conversion.
- **Handling Multicollinearity:** The VIF analysis helped in identifying and addressing multicollinearity issues, ensuring that the model predictions were reliable.
- **Optimized Threshold:** The case study emphasizes the need to optimize the threshold for classification rather than relying on the default 0.5. This was likely done through ROC curve analysis and selecting the cutoff that maximizes both sensitivity and specificity.**Model**
- **Performance:** The final model had a good balance between precision and recall, indicating that it could effectively distinguish between leads likely to convert and those that are not.



# CONCLUSION

---

1. **Effective Model Performance:** The lead scoring model developed shows strong predictive capabilities with consistent performance across both the train and test sets, achieving an accuracy of approximately 79% on both datasets.
  2. **Balanced Sensitivity and Specificity:** The model demonstrates a good balance between sensitivity and specificity, ensuring that it accurately identifies both potential converters and non-converters, minimizing the risk of false positives and false negatives.
  3. **Optimized Lead Scoring:** By identifying a threshold of 47%, the model effectively distinguishes "Hot Leads" with a high likelihood of conversion, allowing the sales team to focus their efforts on the most promising prospects.
  4. **High Precision and Recall:** The precision and recall scores indicate that the model reliably predicts lead conversions, contributing to a more efficient lead conversion process with fewer wasted resources.
  5. **Future-Proofing Considerations:** The model is designed with flexibility in mind, capable of adapting to future constraints such as peak times and optimal manpower utilization, ensuring sustained performance even as business conditions evolve.
  6. **Strategic Impact:** Implementing this lead scoring model is expected to significantly enhance the lead conversion rate, aligning with the CEO's target of an 80% conversion rate and driving higher revenue through targeted sales efforts.
- This conclusion highlights the key achievements and strategic benefits of the lead scoring model, positioning it as a valuable tool for improving sales efficiency and conversion rates