

Lead Score Analysis

Logistic Regression Model Analysis for Lead Conversion

- Santhosh Balaji
 - Jyoti Singh
 - Bhanu Teja
 - Deep Sheth

Objective of the Analysis

- **Key Point:**

Analyze factors contributing to lead conversion and optimize efforts for better conversion rates.

- **Objective:**

- Identify significant features influencing conversion.
- Develop and validate a logistic regression model.
- Evaluate performance with accuracy, sensitivity, and specificity.

Data Overview

- **Dataset Details:**

- **Total Observations:** 4,572

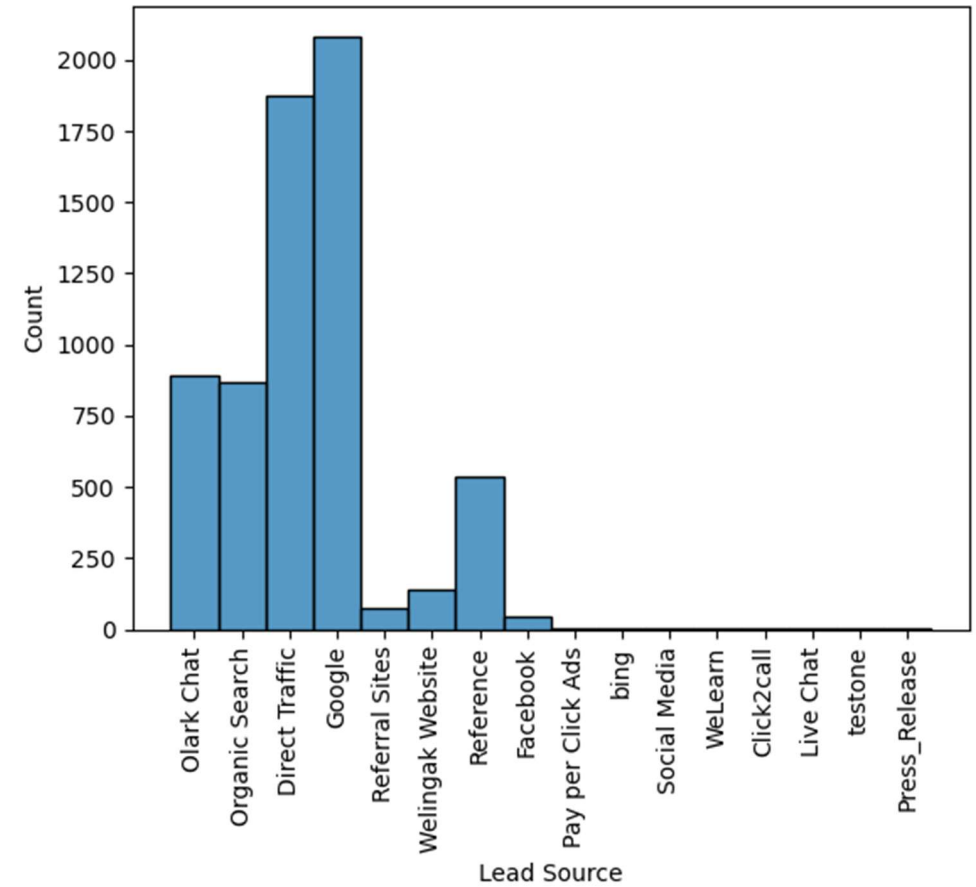
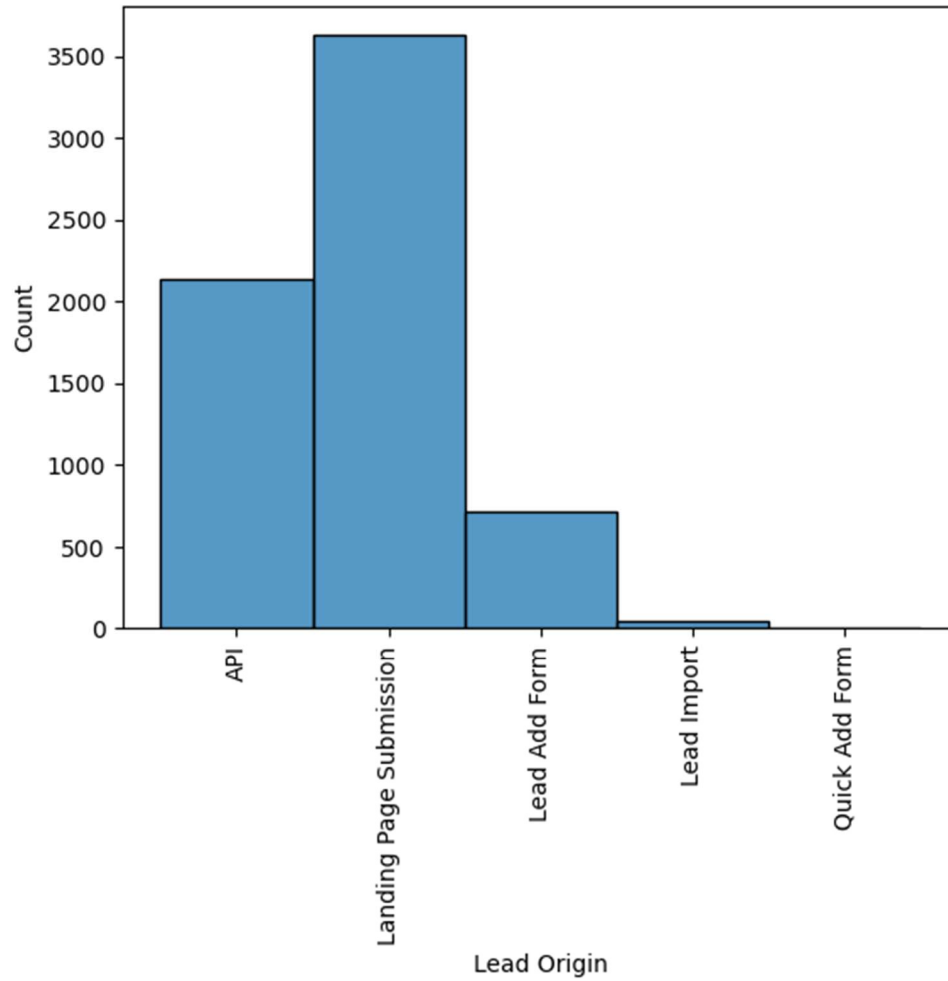
- **Key Features:**

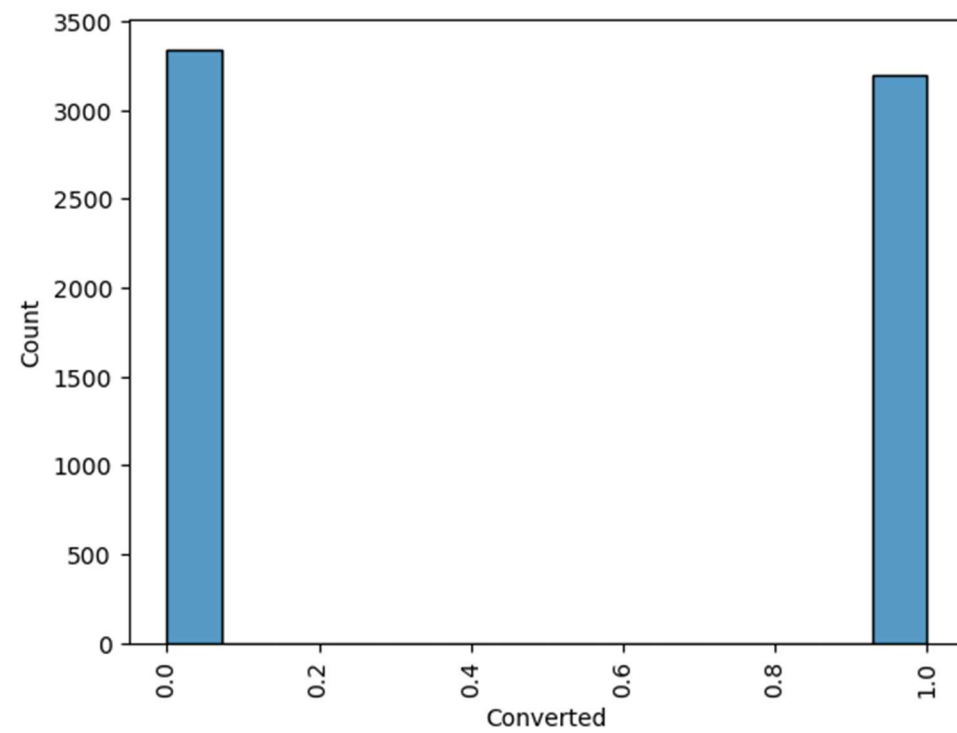
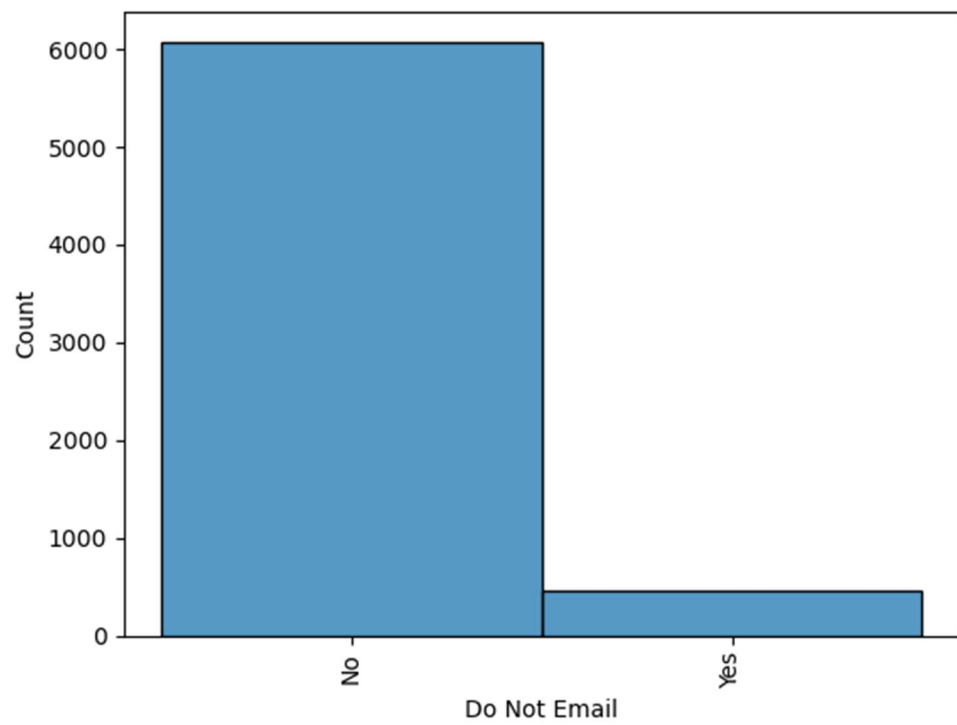
- Total Visits
- Total Time Spent on Website
- Lead Origin
- Lead Source
- Last Notable Activity

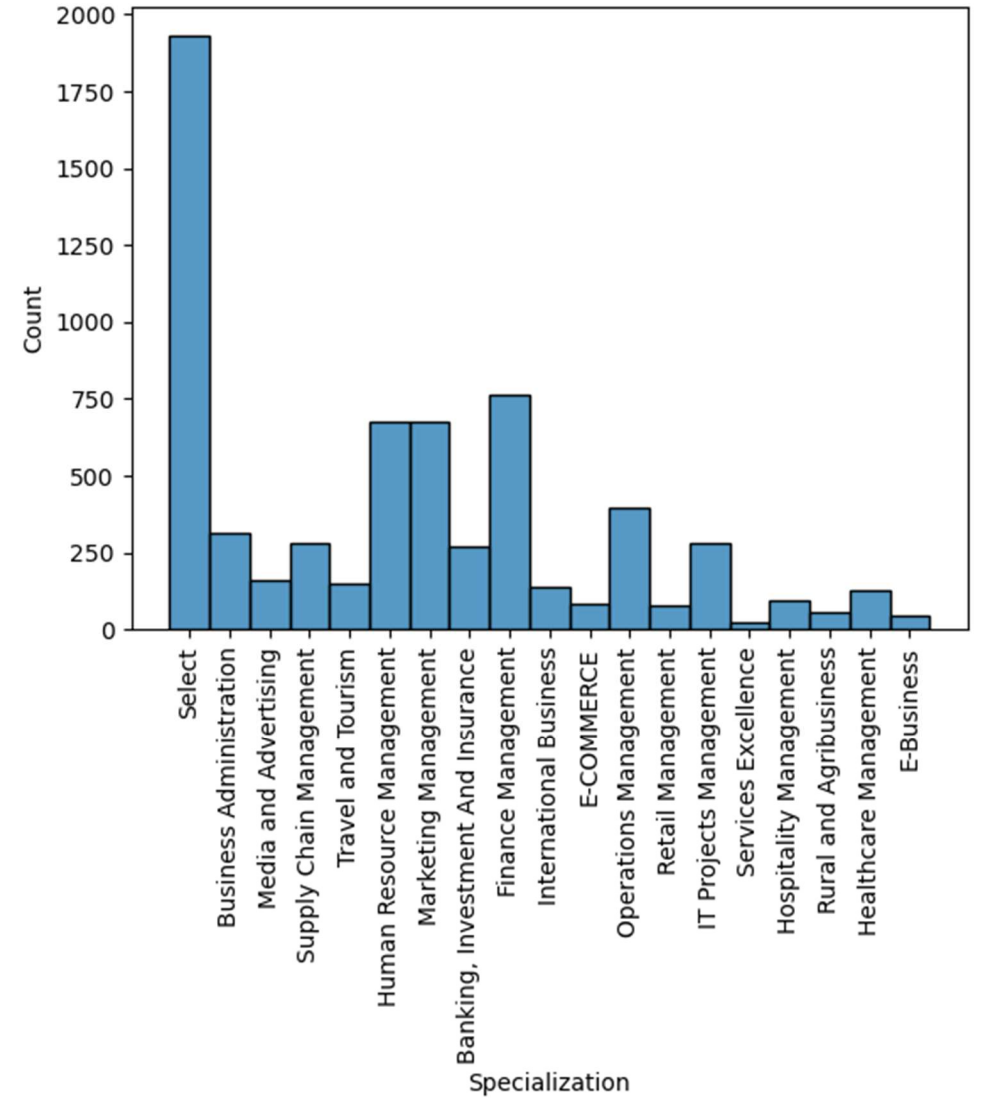
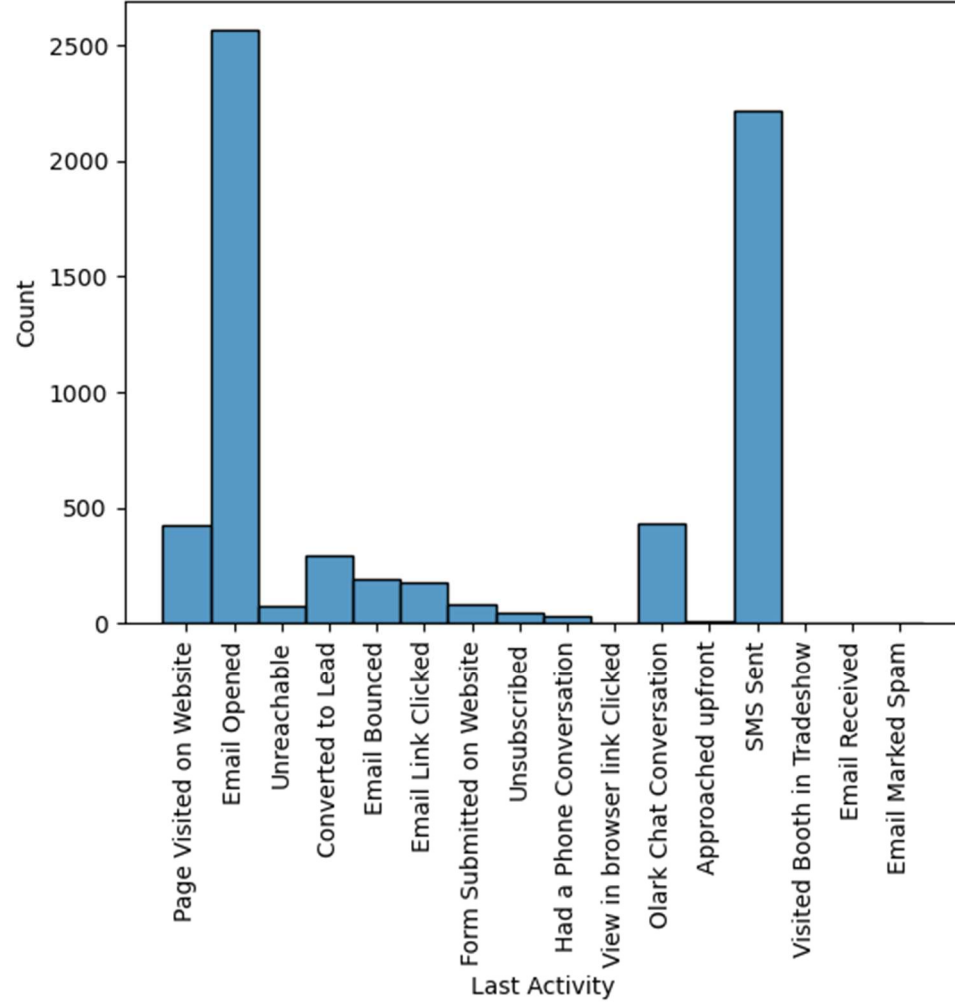
- **Missing Values:**

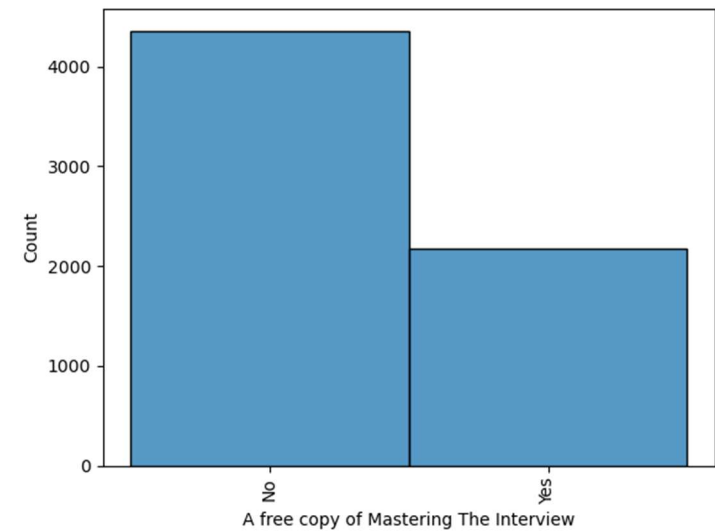
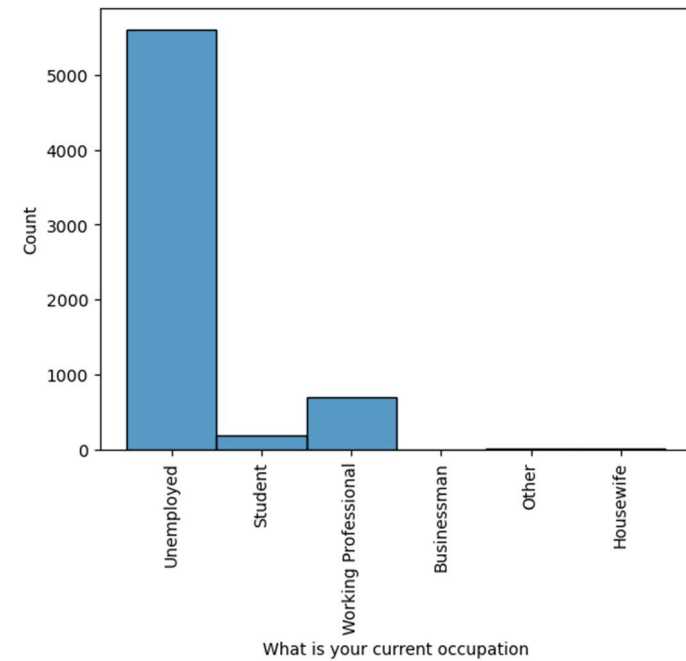
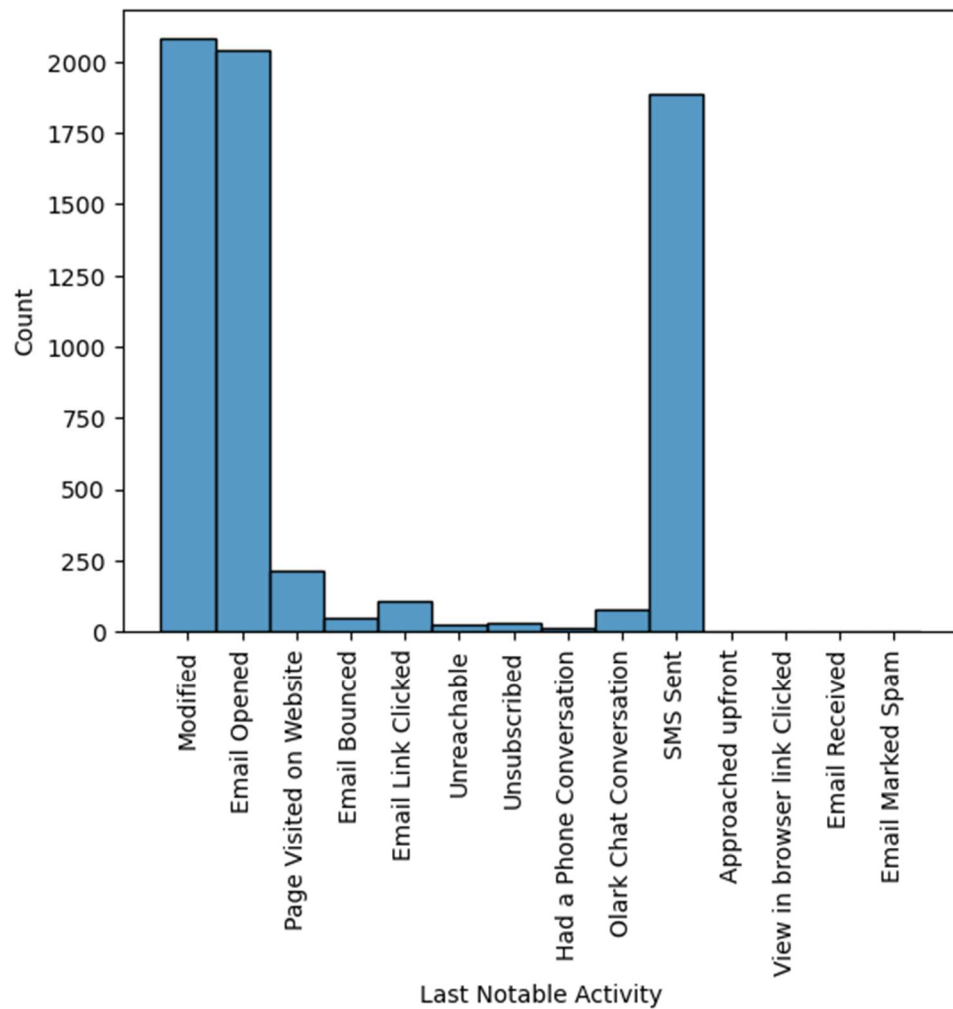
- Several features with null values (Lead Quality, Profile Scores, etc.).

Missing value percentages (Bar Chart)

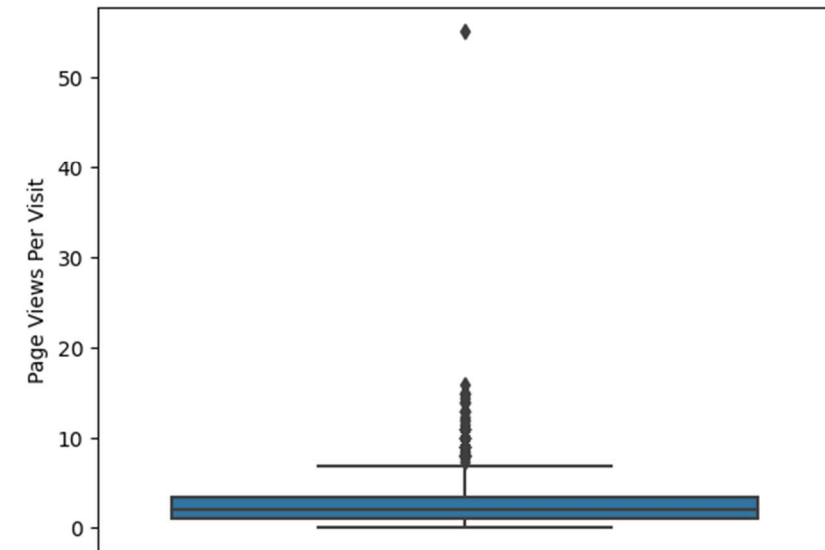
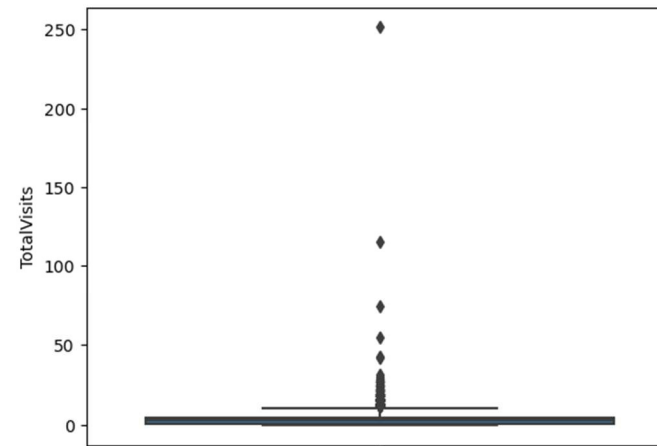
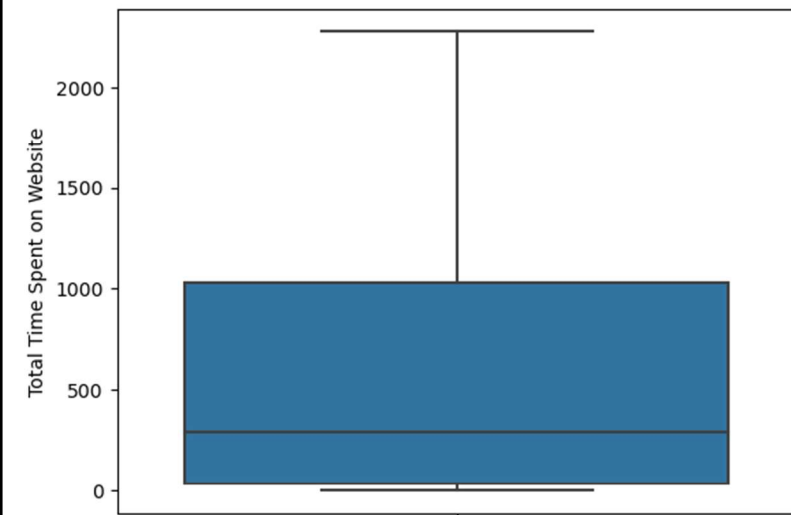








Boxplots for TotalVisits, Page Views Per Visit, and Total Time Spent on Website



Feature Selection and Importance

- Top 3 Features (Continuous):**

- 1.TotalVisits
- 2.Total Time Spent on Website
- 3.Page Views Per Visit

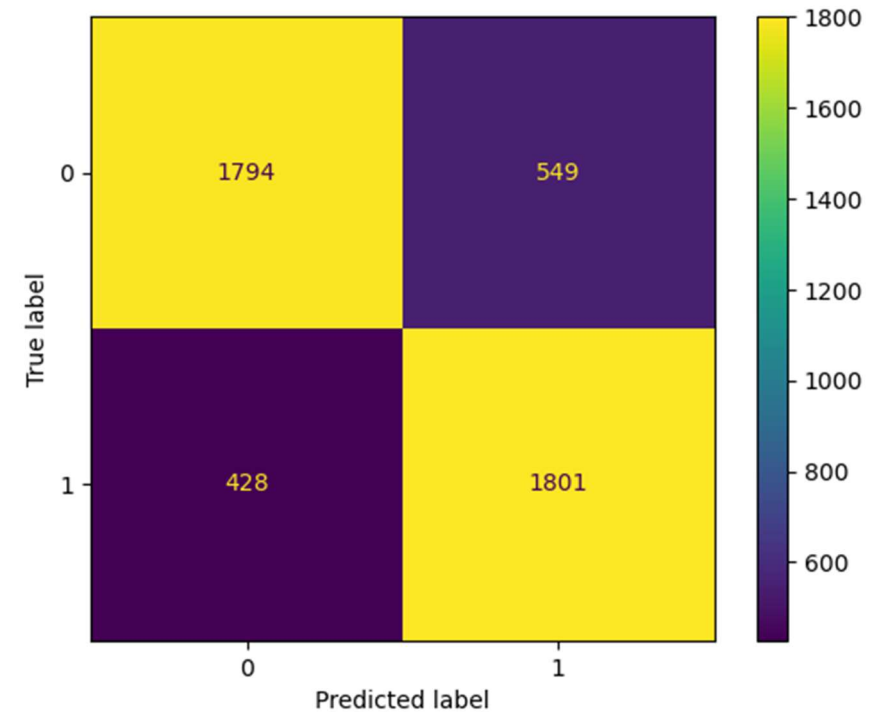
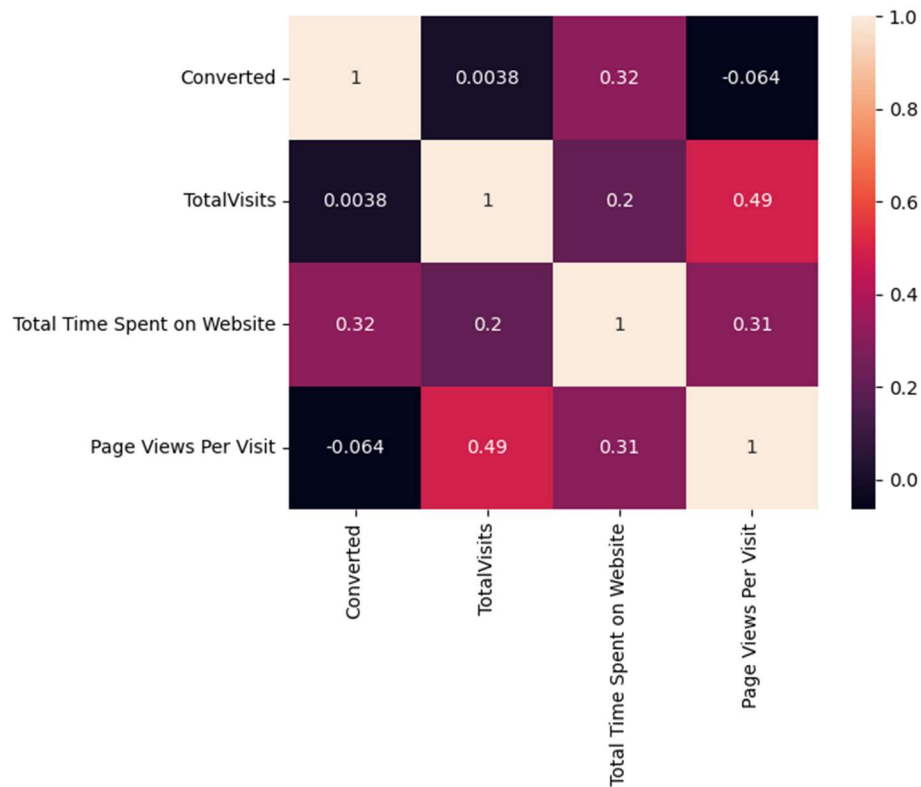
- Top Categorical Features:**

- 1.Lead Source - Welingak Website
- 2.Lead Origin - Lead Add Form
- 3.Last Notable Activity - Unreachable

Logistic Regression Model

- **Model Selection:** Logistic Regression
- **Variables and Coefficients:**
 - TotalVisits: **10.48**
 - Total Time Spent on Website: **4.39**
 - Lead Source - Welingak Website: **2.92**
 - Last Notable Activity - Unreachable: **2.85**
- **Multicollinearity Check:**
 - Variance Inflation Factor (VIF) < 5 for all variables.

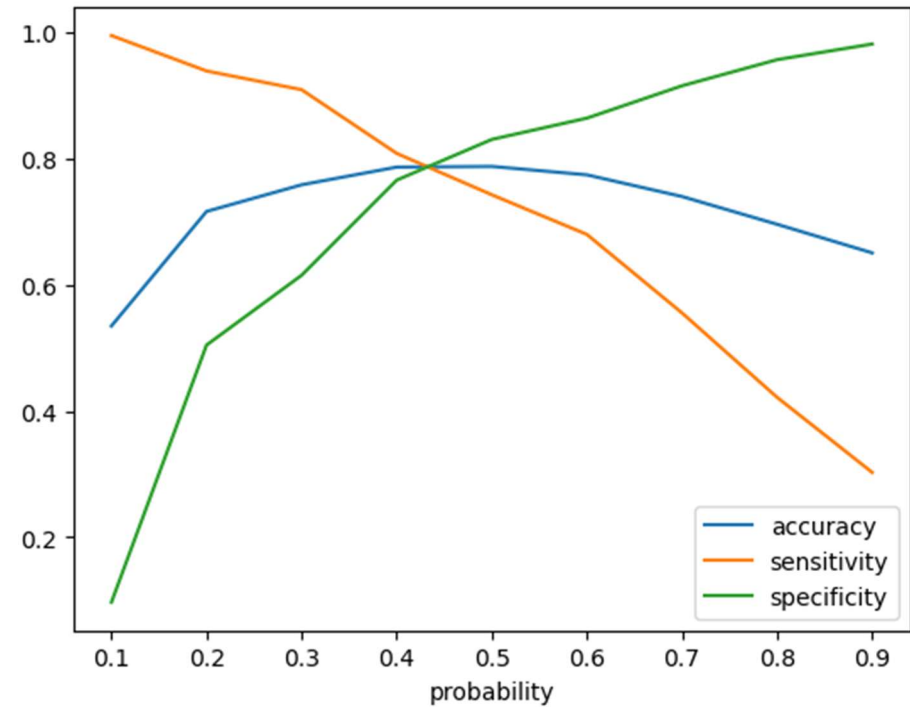
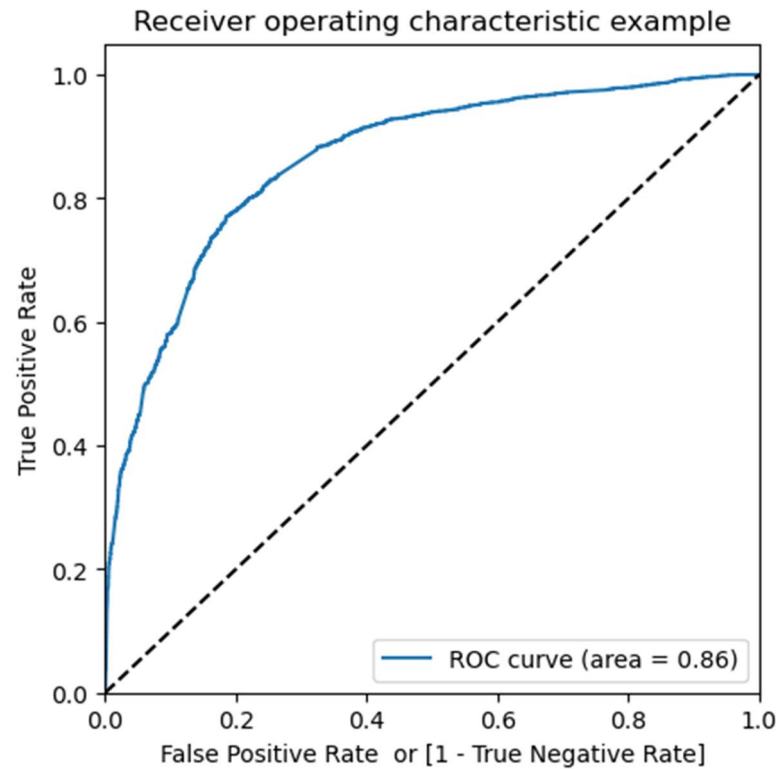
Correlation Heatmap & Confusion Matrix



Metrics and Cutoff

- **Optimal Probability Cutoff:** 0.44
- **Evaluation Metrics at Cutoff:**
- **Accuracy:** ~76.6%
- **Sensitivity (Recall):** Ensures more positives are captured.
- **Specificity:** Balances false positives and false negatives.

ROC Curve & Confusion Matrix



Key Insights

1.Visits Matter:

More website visits increase the likelihood of conversion.

2.Time on Website:

The time spent correlates significantly with lead conversions.

3.Top Sources:

Focus on **Welingak Website** and **Lead Add Form** for better engagement.

Recommendations

- Increase user engagement by optimizing website experience.
- Target leads that spend more time and revisit the website.
- Focus marketing efforts on **Lead Add Form** and **Welingak Website** sources.
- Monitor "Unreachable" leads for additional follow-ups.

Conclusion

- **Final Model Status:**

- All variables significant ($p\text{-value} < 0.05$).
- No multicollinearity.

- **Next Steps:**

- Deploy the model for real-time prediction.
- Optimize probability cutoffs for specific business goals.