Lead Scoring Case Study

Pooja Joshi

Slide 2: Problem Statement

Challenge: Low lead conversion rate (~30%).
Objective: Develop a logistic regression model to predict lead conversion probability and improve conversion rates to ~80%.
Goal: Assign lead scores to prioritize “Hot Leads” for the sales team.

Diagram: Funnel Illustration

- **Stages**: Leads (10,000) → Qualified Leads (5,000) → Converted Leads (3,000).

Slide 3: Approach

Data Preprocessing:

- Cleaning null values and irrelevant levels (e.g., 'Select').
- Encoding categorical variables.

Exploratory Data Analysis (EDA):

- Identified top-performing features through visualizations and statistical tests.

Model Building:

- Logistic Regression with Recursive Feature Elimination (RFE) to select top predictors.

Evaluation:

- Metrics: Accuracy: **79.6%**, ROC-AUC: **0.83**, Confusion Matrix: (see Slide 5).

Recommendations:

- Developed actionable insights for better lead conversion strategies.

Diagram: Workflow diagram: Preprocessing → EDA → Model Building → Evaluation → Recommendations.

Slide 4: Key Insights from Data

Top Predictors:

- Total Time Spent on Website (45% importance).
- Lead Source (35% importance, e.g., Google, Organic Search).
- Last Activity (20% importance, e.g., Email Opened, Form Submitted).

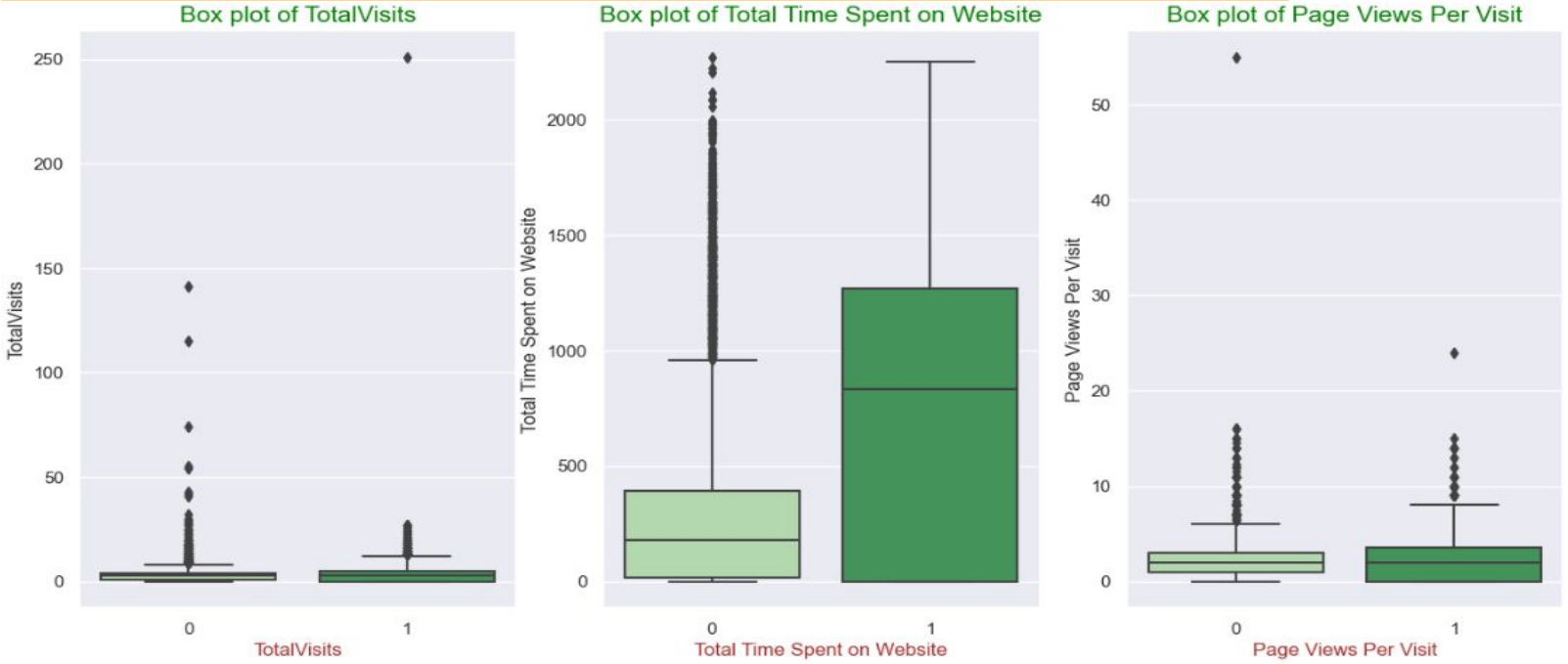
Conversion Trends:

- High engagement (time spent on the website) correlates strongly with lead conversion.
- Specific sources like Google drive high-quality leads.

Diagram: Bar chart of feature importance.

Diagram:

- Box plots of *Total Visits*, *Total Time Spent on Website*, and *Page Views Per Visit*, showing engagement variability among leads.



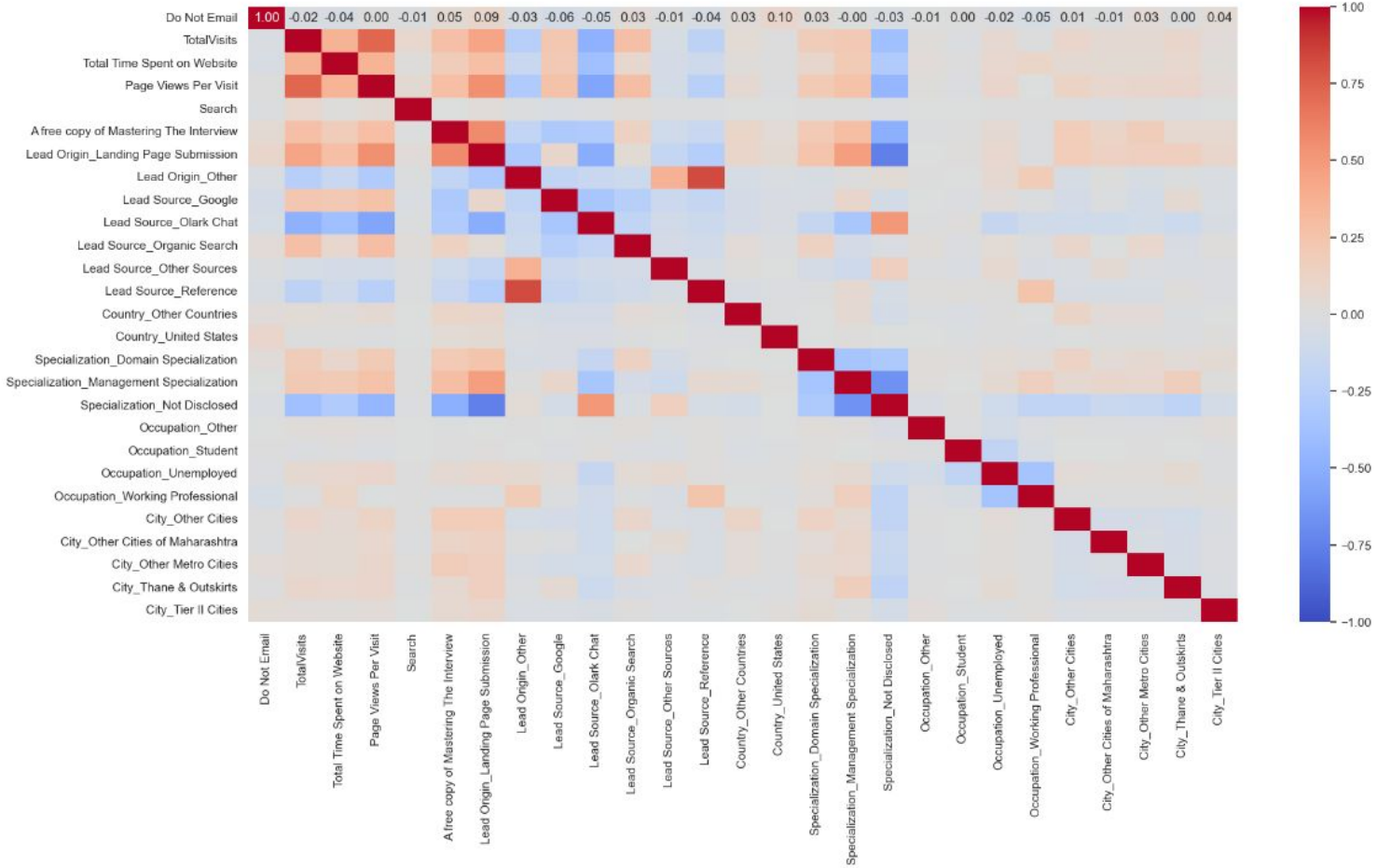
Model Results

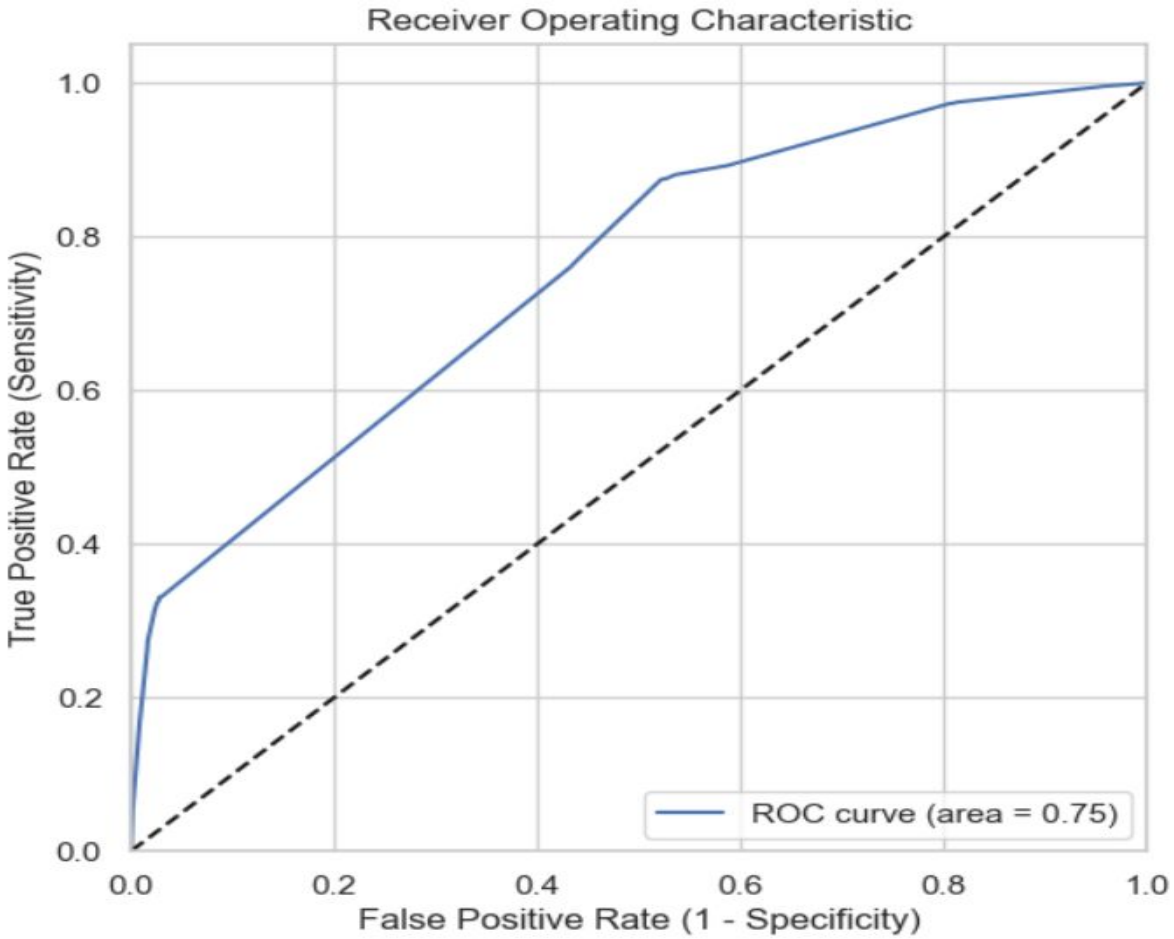
Performance Metrics:

- **Accuracy**: 79.6%.
- **ROC-AUC**: 0.83.
- **Confusion Matrix**:
  - True Positives: 3,000.
  - False Positives: 500.
  - Precision: 85%.
  - Recall: 78%.

Diagrams:

- ROC Curve for model performance.
- Heatmap for confusion matrix.


Do Not Email
TotalVisits
Total Time Spent on Website
Page Views Per Visit
Search
A free copy of Mastering The Interview
Lead Origin_Landing Page Submission
Lead Origin_Other
Lead Source_Google
Lead Source_Olark Chat
Lead Source_Organic Search
Lead Source_Other Sources
Lead Source_Reference
Country_Other Countries
Country_United States
Specialization_Domain Specialization
Specialization_Management Specialization
Specialization_Not Disclosed
Occupation_Other
Occupation_Student
Occupation_Unemployed
Occupation_Working Professional
City_Other Cities
City_Other Cities of Maharashtra
City_Other Metro Cities
City_Thane & Outskirts
City_Tier II Cities
1.00 -0.02 -0.04 0.00 -0.01 0.05 0.09 -0.03 -0.06 -0.05 0.03 -0.01 -0.04 0.03 0.10 0.03 -0.00 -0.03 -0.01 0.00 -0.02 -0.05 0.01 -0.01 0.03 0.00 0.04
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Receiver Operating Characteristic
True Positive Rate (Sensitivity)
False Positive Rate (1 - Specificity)
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.75)

Lead Scoring

Definition: Assigning a score (0-100) based on conversion likelihood.

Implementation:

- Leads with scores >80 are prioritized as “Hot Leads.”

Impact:

- Focuses the sales team on high-potential leads, improving efficiency.

Diagram: Histogram of lead score distribution.

Strategies for Different Scenarios

Internship Phase:

- Focus on “Hot Leads.”
- Rapid follow-ups for high-scoring leads.
- Real-time monitoring of activities.

Minimizing Calls:

- Adjust lead score threshold to >90.
- Automate communication for lower-priority leads.
- Use activity-based prioritization (e.g., time spent on the website).

Recommendations

Enhance Website Engagement:

- Improve content and navigation.
- Use chatbots to guide potential leads.

Optimize Marketing Channels:

- Invest in Google and Organic Search campaigns.

Train Interns and Employees:

- Focus on interpreting lead scores and prioritizing follow-ups.

Automate Lower-Priority Tasks:

- Use email campaigns for low-probability leads.

Conclusion

Summary:

- The lead scoring model is an effective tool for identifying high-potential leads.
- Aligning sales and marketing efforts using the model improves efficiency and conversion rates.

Future Work:

- Continuous refinement of the model and integration of additional data sources.