Lead Scoring Case Study

Lead scoring model to improve conversion rates.

by: Ritesh Ranjan, Raseena Ameer & Reema David

Business Challenge

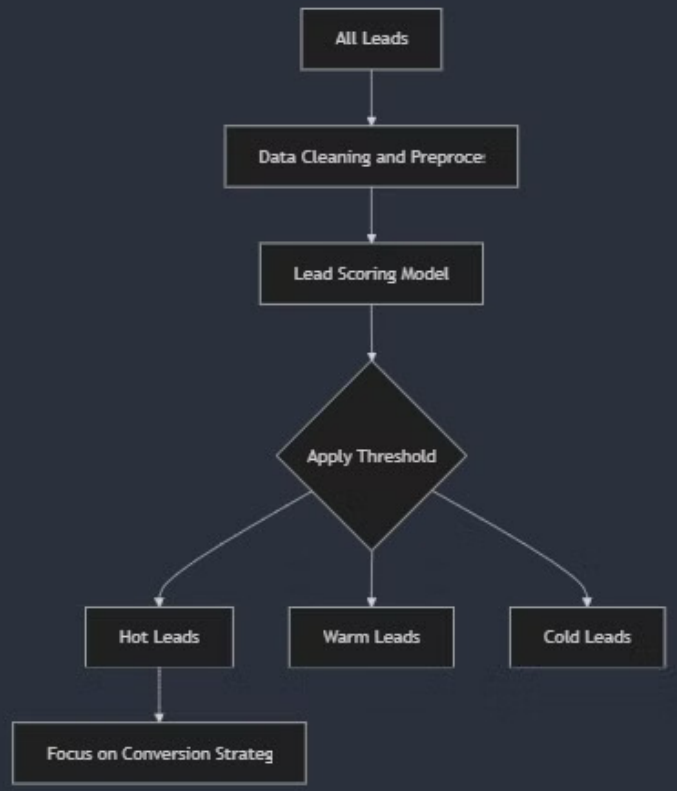
Our lead conversion rate is only 38.5%, far below our 80% target. This indicates inefficiencies in lead nurturing, wasting resources and opportunities. A lead scoring model will help.

Prioritize Leads

A lead scoring model helps prioritize high-potential leads, optimizing resource allocation and maximizing ROI.

Improve Efficiency

Efficiently identify and engage potential customers by focusing on qualified leads, improving sales efficiency and reducing wasted effort.



Objective of the Analysis

Build a predictive model to score leads based on their likelihood of conversion.

1

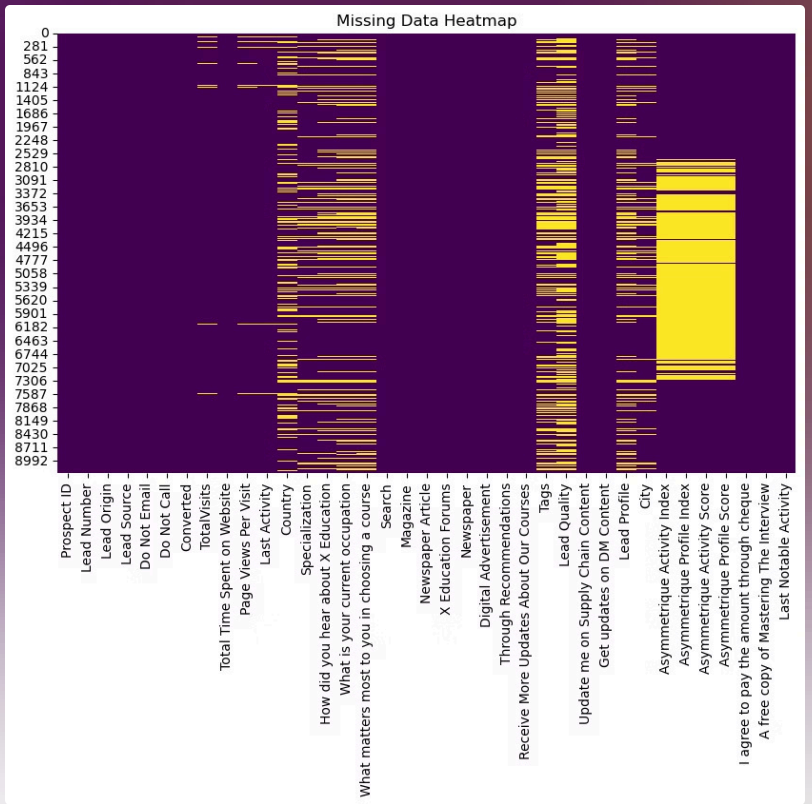
Logistic Regression

Predict lead conversion.

2

Insights

Improve marketing and sales focus.



Data Exploration and Preprocessing

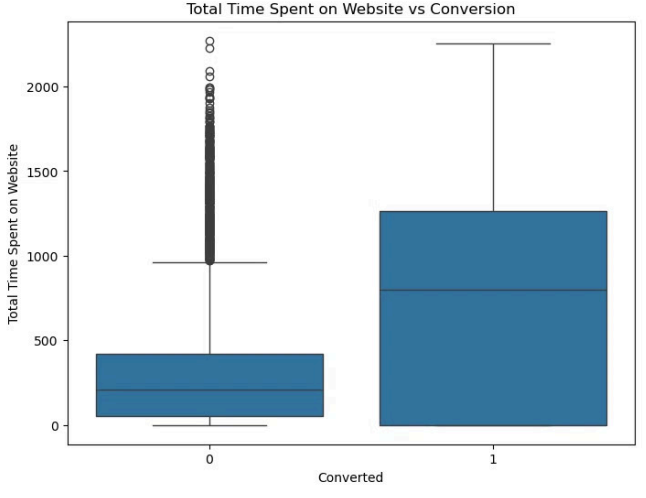
- **Steps in Data Cleaning:**
  a. Addressed missing values: Dropped columns with >30% missing data; imputed or dropped rows where necessary.
  b. Removed "Select" levels in categorical columns like *"Lead Profile"* and *"How did you hear about X Education"*.
  c. Dropped irrelevant features like *"Prospect ID"* and *"Lead Number"*.
  d. Removed duplicate rows and columns.
  e. Created dummy variables for categorical features.
- **Feature Engineering:**
  - Scaled numerical columns using MinMaxScaler.
  - Created derived features where necessary.
- **Visualization:**
  - Include bar charts for missing values.
  - Add a bar chart for unique values in categorical columns.

Exploratory Data Analysis (EDA)

Insights from exploratory data analysis.

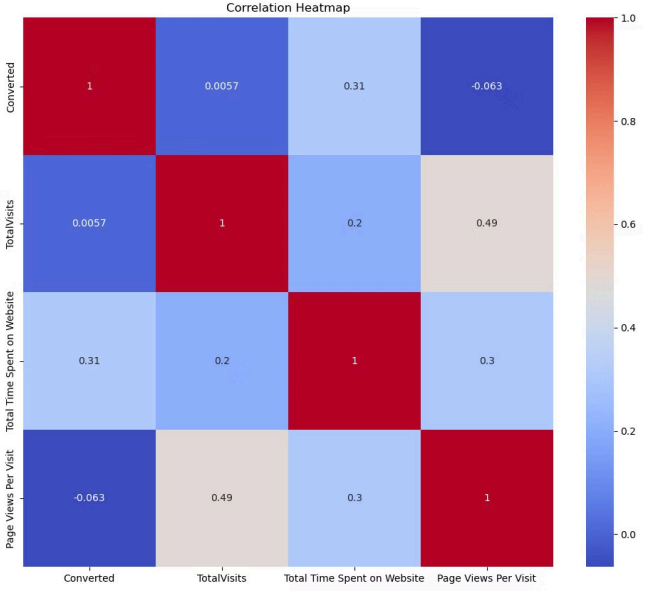
Conversion Rate Distribution

Bar chart of the "Converted" column.



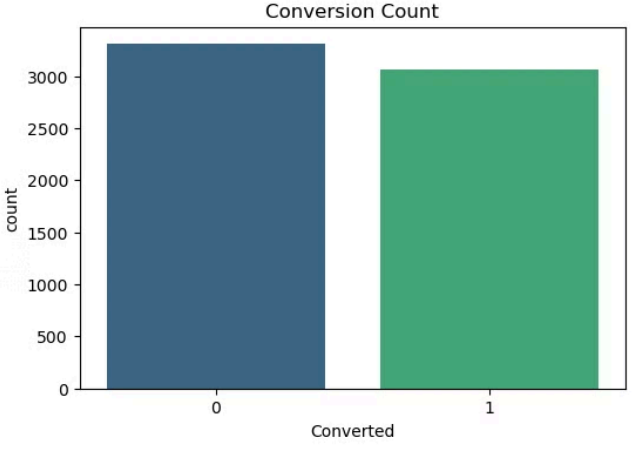
Numerical Feature Relationships

Box plots and heatmap.



Categorical Feature Insights

Conversion rates across top categorical variables.



Model Development

Logistic regression model for lead conversion.

1 Target Variable

"Converted" (binary classification).

2 Split Dataset

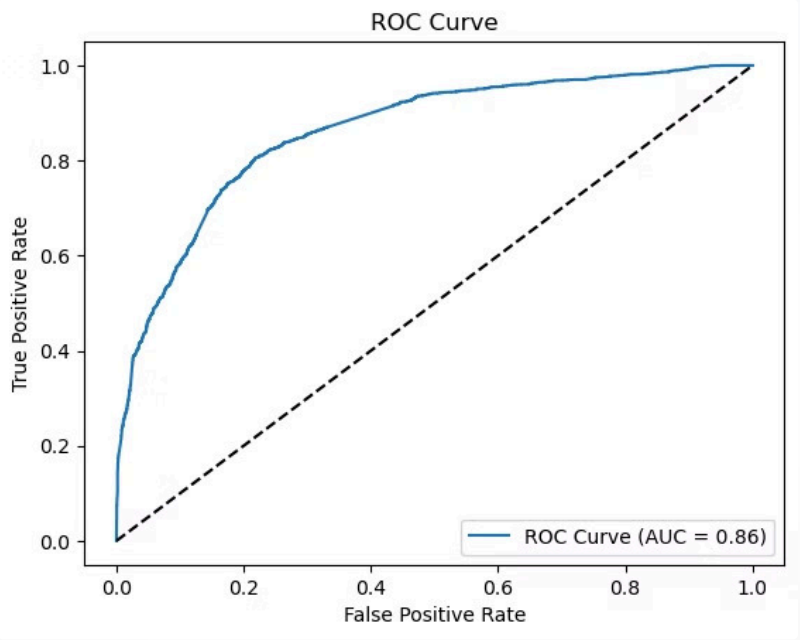
70% train and 30% test.

3 Feature Selection

Feature selection using Recursive Feature Elimination (RFE): Retained 15 important features.

4 Evaluation Metrics

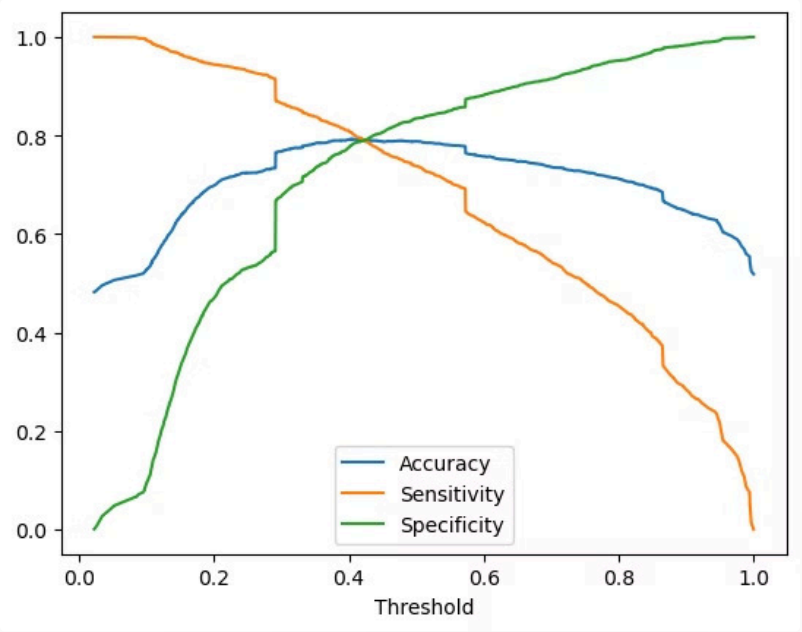
Precision, Recall, Sensitivity, Specificity, ROC Curve, and AUC score.



Model Evaluation

Performance metrics and drivers of lead conversions.

	Training	Testing
Sensitivity	0.73	0.73
Specificity	0.83	0.83
Accuracy	0.78	0.78



Drivers of Lead Conversion

- **Top Features:**
  - **Numerical:** *"Total Time Spent on Website"*, *"Total Visits"*, etc.
  - **Categorical:** *"Lead Source"*, *"Last Activity"*, etc.
- **Recommendations for Lead Prioritization:**
  - During **aggressive periods**: Lower the threshold to focus on hot leads.
  - During **low-priority periods**: Focus on high-probability leads to maximize efficiency.

Recommendations and Action Plan

Recommendations and action plan for lead scoring.

Recommendations:

- Prioritize leads based on predicted scores to improve conversion rates.
- Optimize marketing spend by focusing on high-conversion lead sources.
- Regularly monitor model performance and recalibrate thresholds as necessary.

Action Plan:

- Implement the logistic regression model in the sales funnel.
- Train the sales team on how to use lead scores.
- Continuously refine the model with new data.

Conclusion

The **Lead Scoring Case Study** aimed to identify key drivers of lead conversion and develop a predictive model to improve conversion rates. By applying **data preprocessing techniques** such as handling missing values, encoding categorical variables, and removing irrelevant features, we ensured a clean dataset for model training.

Through **exploratory data analysis (EDA)**, we uncovered critical insights into feature relationships and their impact on conversion. Key features like **'Total Time Spent on Website'** and **'Lead Source'** emerged as significant contributors to lead conversion, highlighting the importance of focused marketing efforts.

We built a **logistic regression model** to predict lead conversion, achieving a strong performance with a sensitivity of **73.8%** and specificity of **83.5%**. The model allowed us to effectively classify leads into categories such as **hot**, **warm**, and **cold**, enabling targeted follow-up actions to prioritize high-conversion potential leads.

Additionally, we used **recursive feature elimination (RFE)** to select the most impactful features and further fine-tuned the model. The **ROC curve** and other performance metrics reinforced the model's effectiveness in distinguishing between positive and negative outcomes.