Lead Scoring Case Study: Data Analysis and Model Building

1. Introduction

- The objective of this analysis is to build a logistic regression model to assign lead scores between 0 and 100 to potential leads, aiding the company in targeting potential customers effectively.

2. Data Overview

- The dataset comprises various features such as lead origin, lead source, total visits, total time spent on the website, and more.
- Initial exploration revealed missing values in several columns, which were handled by imputation or replacement.

3. Data Preprocessing

- Irrelevant columns ('Prospect ID', 'Lead Number') were dropped from the dataset.
- Categorical variables with 'Select' as a category were replaced with NaN values.
- Missing values in numerical variables were filled with mean values, while categorical variables were imputed with 'Unknown'.
- One-hot encoding was applied to categorical variables for model compatibility.

4. Model Building

- The dataset was split into training and testing sets with an 80-20 split.
- Features were scaled using StandardScaler to ensure uniformity.
- Hyperparameter tuning was performed using GridSearchCV to find the best parameters for the logistic regression model.
- The model achieved an accuracy of 94.06% on the training set.

5. Model Evaluation

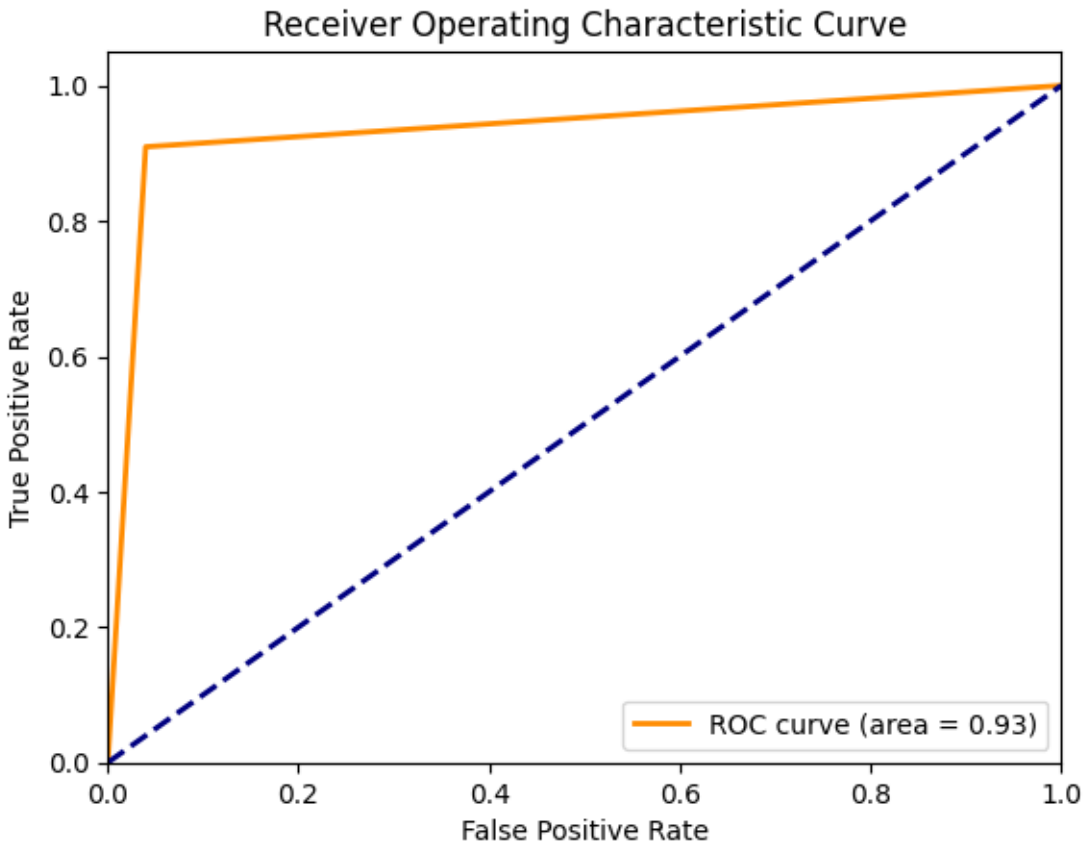
- The model was evaluated on the testing set using metrics such as confusion matrix and classification report.
- Confusion Matrix:

```
[[1063 44]
 [ 67 674]]
```

	precision	recall	f1-score	support
0	0.94	0.96	0.95	1107
1	0.94	0.91	0.92	741

accuracy			0.94	1848
macro avg	0.94	0.93	0.94	1848
weighted avg	0.94	0.94	0.94	1848

- The ROC curve demonstrates the model's performance with an area under the curve (AUC) of 0.94.



6. Recommendations

- The logistic regression model shows promising results in predicting lead conversion.
- Further analysis could focus on feature engineering and exploring additional models to improve performance.
- Regular model monitoring and updates are essential to adapt to changing business requirements.

7. Conclusion

- In conclusion, the logistic regression model provides valuable insights into lead scoring, enabling the company to prioritize and target potential customers efficiently.