Lead Scoring Analysis for X Education

This presentation will provide a detailed analysis of the lead scoring model developed for X Education, a leading provider of online educational services. The project aims to understand the data, build a logistic regression model, and evaluate its performance in identifying high-potential leads.

by Sahil Verma

Introduction

1 Overview

This analysis explores the purpose, dataset, and key objectives of the lead scoring project for X Education.

2 Objectives

The main goals are to understand and clean the data, build a logistic regression model, and evaluate its performance.

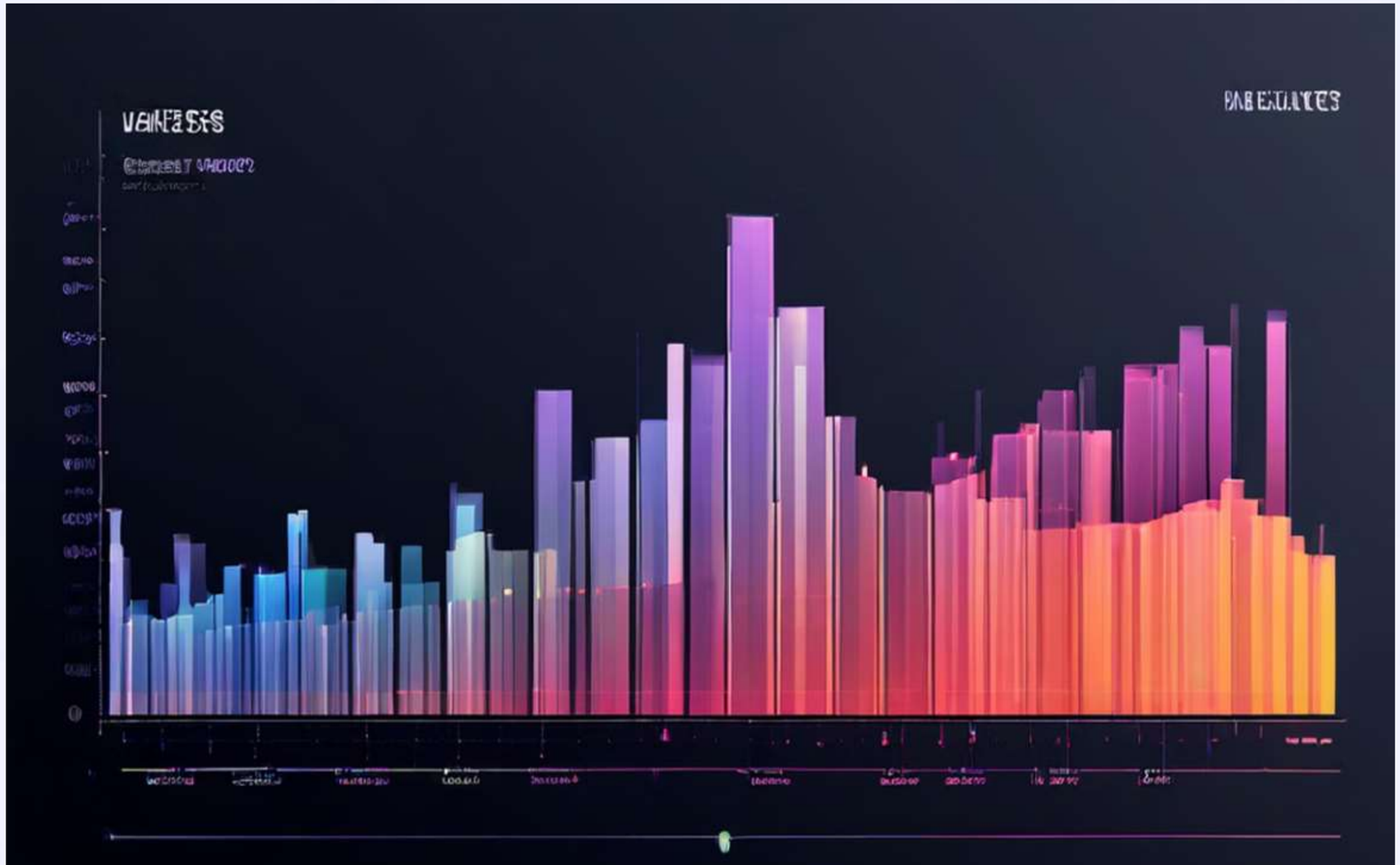
Data Understanding and Cleaning

Data Description

The dataset contains 9,240 entries with 37 features, which requires careful inspection and preprocessing.

Data Inspection

Summary statistics and handling of missing values are crucial steps in the data understanding phase.



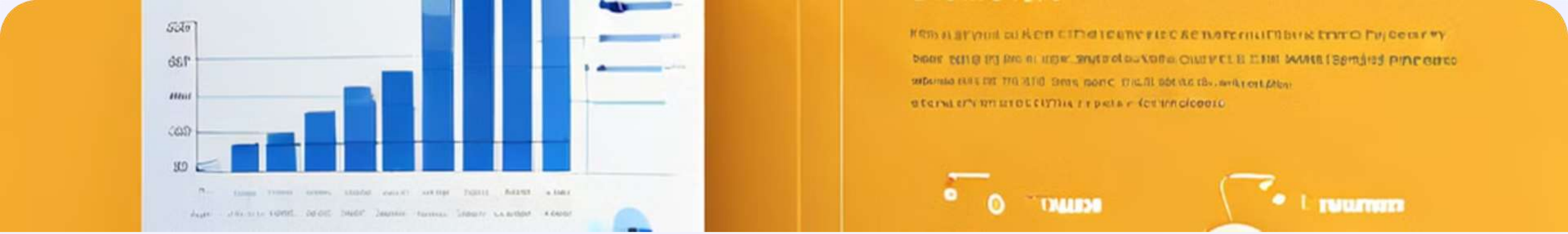
Data Preprocessing

Categorical Variables

Handling categorical variables by converting them to numerical values and dropping unnecessary columns.

Splitting the Data

Dividing the dataset into features (X) and target (y), then splitting it into training and testing sets.



Model Building

Model Selection

Logistic Regression was chosen as the appropriate model for this lead scoring analysis.

1

Training the Model

The logistic regression model was trained on the training dataset, and the summary statistics were analyzed.

2

Correlation Heatmap

A heatmap visualization was created to understand the correlations between features and the target variable.

3

Model Evaluation

1 Evaluation Metrics

The model achieved an accuracy of 77.78%, precision of 79.49%, recall of 93.94%, F1 score of 86.11%, and ROC AUC of 85.35%.

2 ROC Curve

The ROC curve and AUC value were plotted to provide a visual representation of the model's performance.

Insights and Recommendations

Key Insights

The high recall indicates the model is effective at identifying leads that are likely to convert, while the precision shows the proportion of positive identifications that were accurate.

Recommendations

Focus marketing efforts on leads with high predicted probability of conversion, and conduct further analysis to understand features contributing to non-conversion.

Conclusion and Next Steps

1

Conclusion

The logistic regression model has proven to be an effective tool for lead scoring, summarizing the key findings and its overall effectiveness.

2

Next Steps

Further refinement of the model, integration with business processes, and continuous monitoring and updates are recommended as next steps.

Thank You

Thank you for your attention. We are excited to continue working with X Education to further optimize the lead scoring process and drive business success.