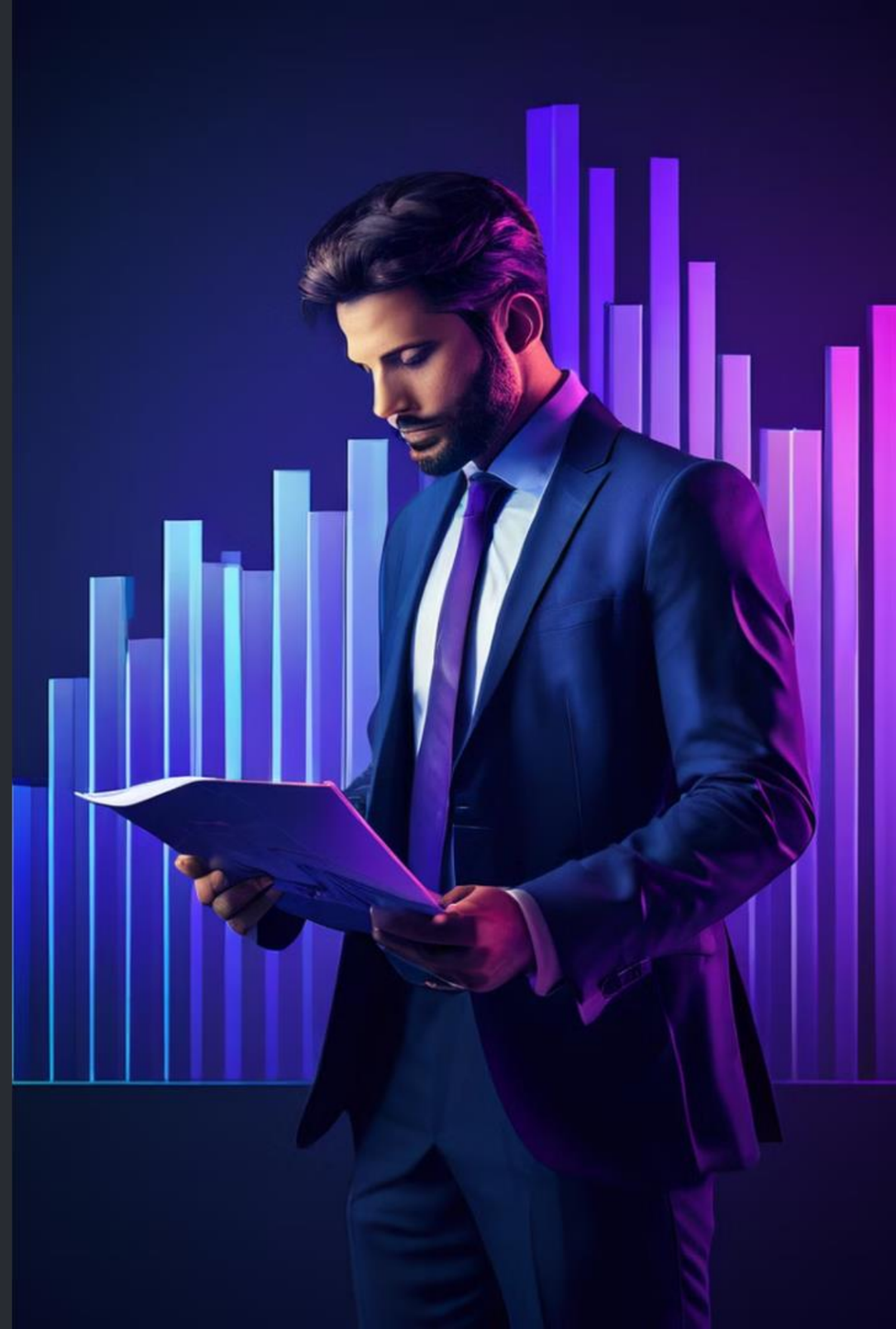
Introduction to Lead Score Case Study

This case study explores how a company developed a lead scoring model to identify high-quality leads and optimize their sales and marketing efforts. The goal was to increase the efficiency of the sales team by focusing on the most promising prospects.

~Prithibi Mondal

Problem Statement

X Education, an online education company, faces a challenge of poor lead conversion despite generating a considerable number of leads daily. The company wishes to enhance its lead conversion rate by identifying the most promising leads, or 'Hot Leads,' who are highly likely to convert into paying customers. To address this, we aim to build a logistic regression model that assigns a lead score between 0 and 100 to each lead based on various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. This lead scoring model will enable X Education to prioritize its efforts towards potential leads, thereby increasing the lead conversion rate and ultimately driving business growth.



Our Approach

1. Data Acquisition and Cleaning
2. Exploratory Data Analysis (EDA)
3. Data Preparation & Feature Scaling
4. Model Development
5. Model Evaluation
6. Accuracy Measurement

Exploratory Data Analysis

Lead Source Analysis

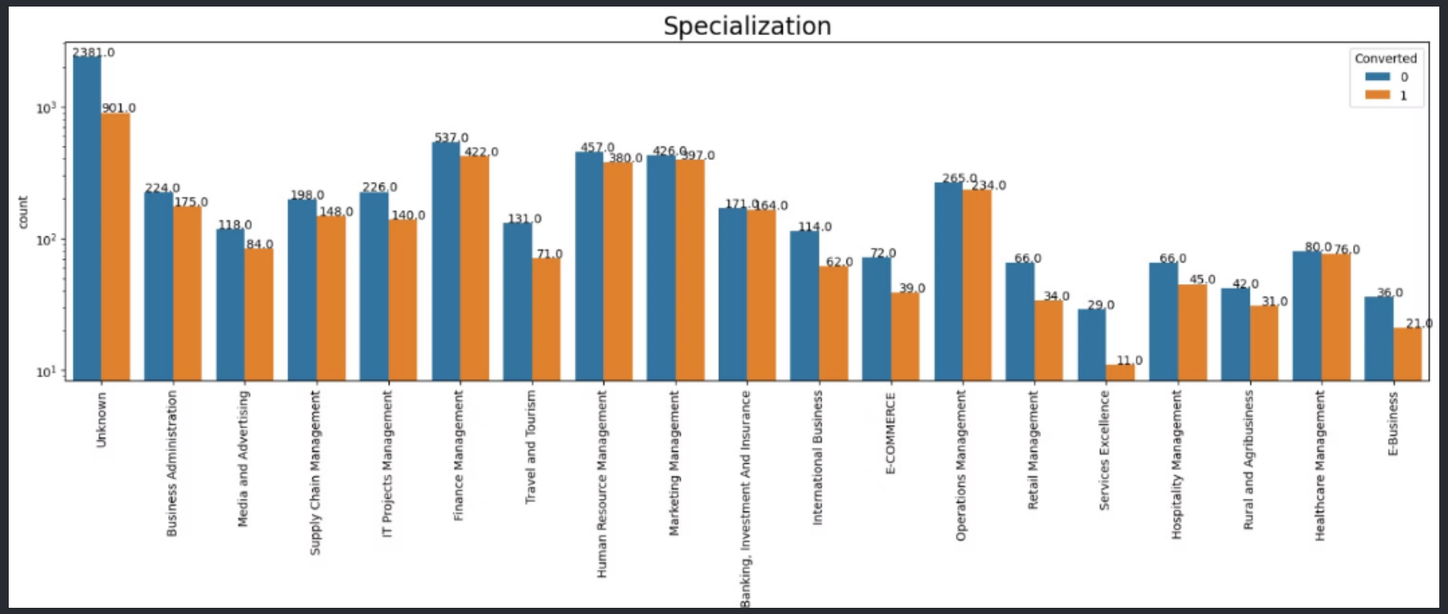
Examined the distribution of leads across different marketing channels to understand their relative effectiveness.

Demographic Factors

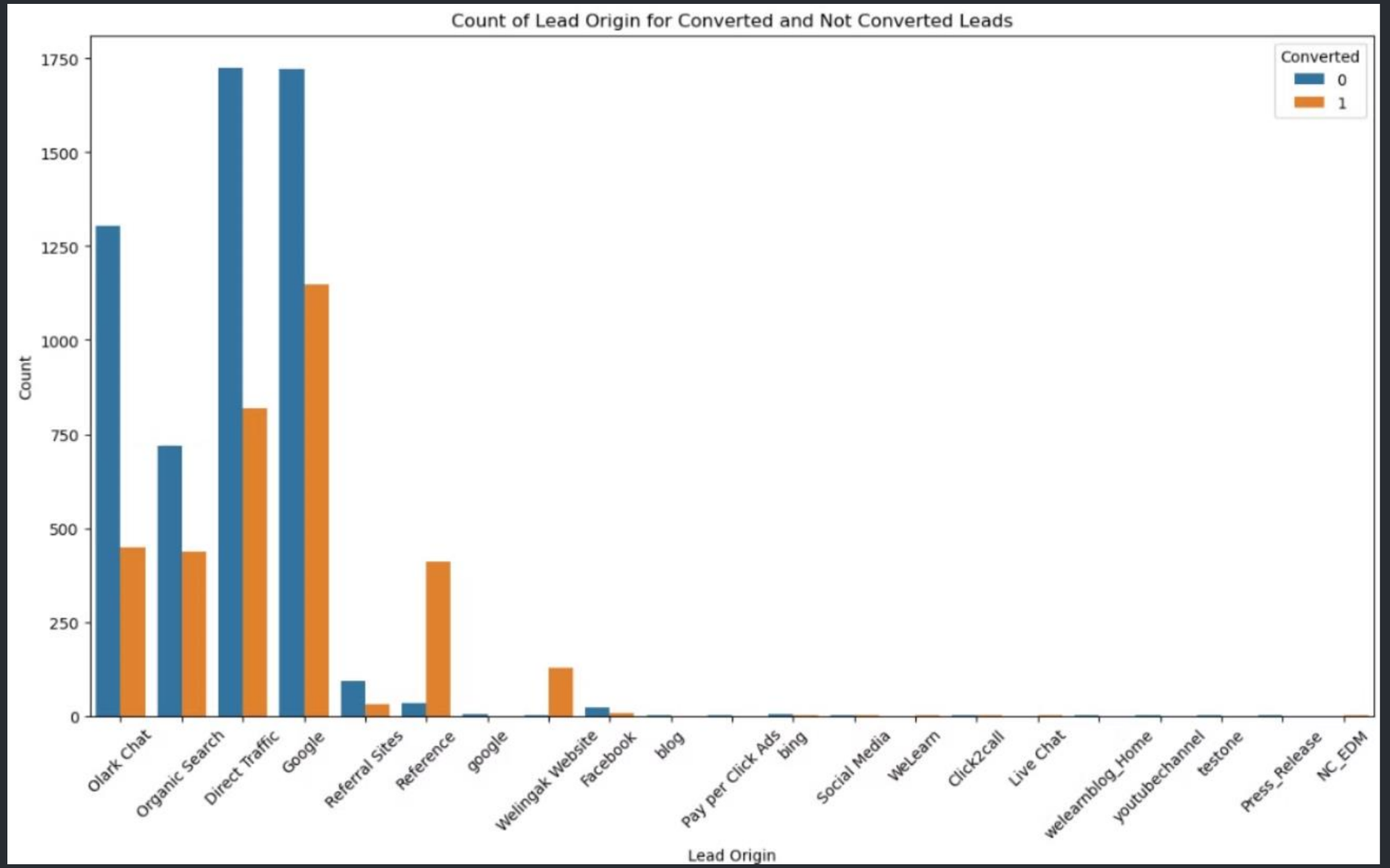
Investigated the impact of factors like company size, industry, and job title on lead conversion rates.

Behavioral Insights

Analyzed lead engagement data, such as website visits, email interactions, and content consumption, to identify patterns and trends.



We can observe from above count plot of 'Specialization' that 'Management' specialization altogether having more number of leads generating. And Other category is also generating more number of leads.



Google and Direct traffic generates maximum number of leads. Conversion rate of 'Reference' and 'Welingak Website' leads is high.

Model Building

1 Feature Selection

- Utilize Recursive Feature Elimination (RFE) to identify the most relevant features for model building.

2 Address Multicollinearity

- Implement Variance Inflation Factor (VIF) analysis to mitigate multicollinearity among predictors.

3 Logistic Regression

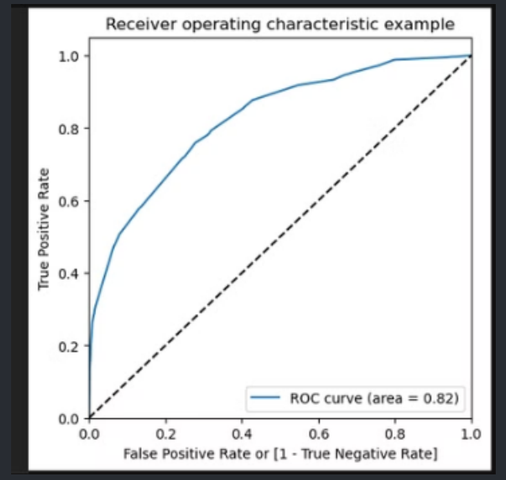
- Construct a logistic regression model using the selected features to predict lead conversion.

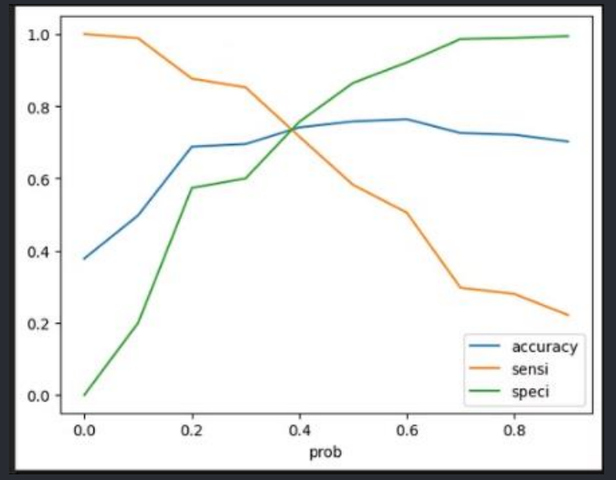
4 Model Evaluation

- Measure the accuracy of the model along with other relevant metrics to gauge its effectiveness in predicting lead conversion.

ROC Curve

Model	AUC	Accuracy	Precision	Recall
Logistic Regression	0.78	0.68	0.78	0.76
Random Forest	0.80	0.81	0.73	0.78





Conclusion and Key Takeaways

In this study, we developed a robust logistic regression model to predict lead conversion for X Education, with a focus on identifying the most promising leads. By leveraging techniques like Recursive Feature Elimination (RFE) and Variance Inflation Factor (VIF) analysis, we selected the most impactful features and addressed multicollinearity. The top three contributing variables to lead conversion were found to be 'Time Spent on Website,' 'Google Search,' and 'Total Visits,' while the key categorical variables to focus on included 'Page Visited on Website,' 'SMS Sent,' and 'Organic Search.'

Prioritizing high-probability leads

To further enhance lead conversion, especially during the aggressive conversion phase with additional intern support.

1

Utilizing automation tools for efficient lead management

To further enhance lead conversion, especially during the aggressive conversion phase with additional intern support.

2

Implementing targeted LinkedIn outreach

To further enhance lead conversion, especially during the aggressive conversion phase with additional intern support.

3

Our model aims to significantly improve the lead conversion rate, aligning with the company's goal of targeting an 80% conversion rate, thereby allowing the sales team to focus their efforts