

Lean Score Case Study

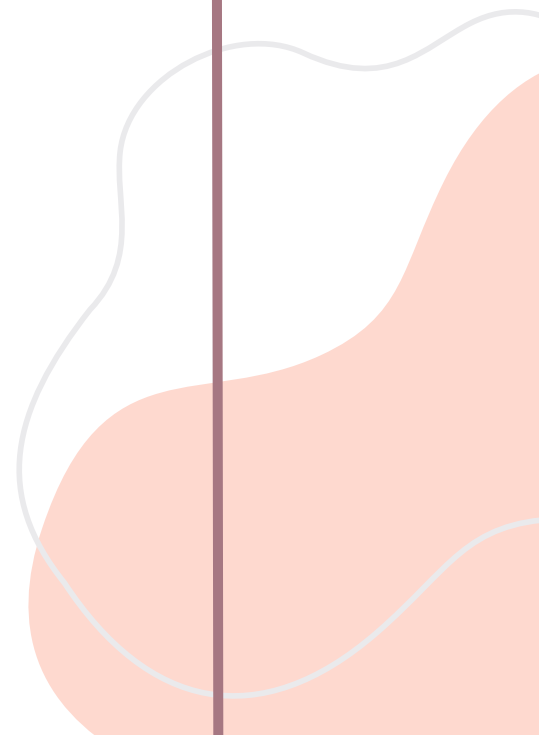
Presented by

- **Kushal Narang**
- **Manasi Deshpande**



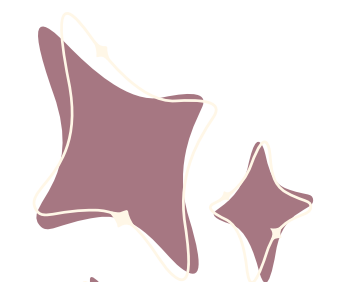
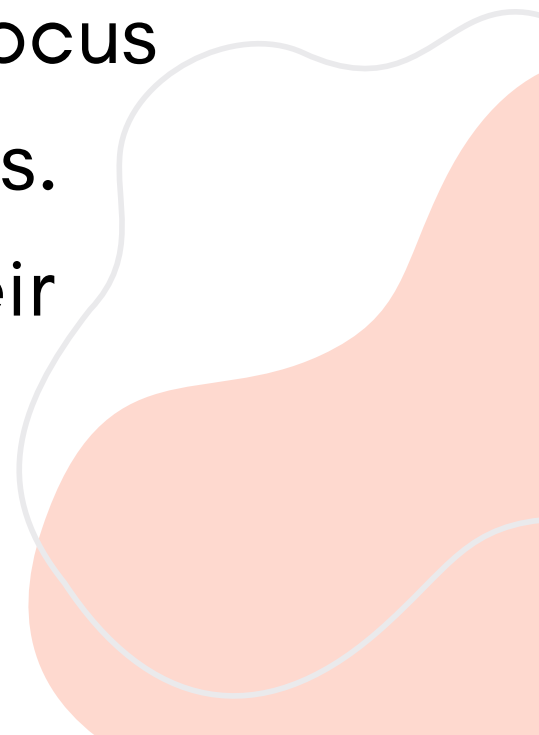


Problem Statement

- X Education provides online courses for industry professionals.
 - Despite acquiring a substantial number of leads, their conversion rate is currently low, at around 30%.
 - To enhance efficiency, the company aims to identify "Hot Leads," or those with the highest potential for conversion.
 - By targeting these promising leads, the sales team can prioritize their efforts and increase the overall conversion rate.
 - This strategic approach will streamline the sales process and maximize the impact of their outreach efforts.
- 



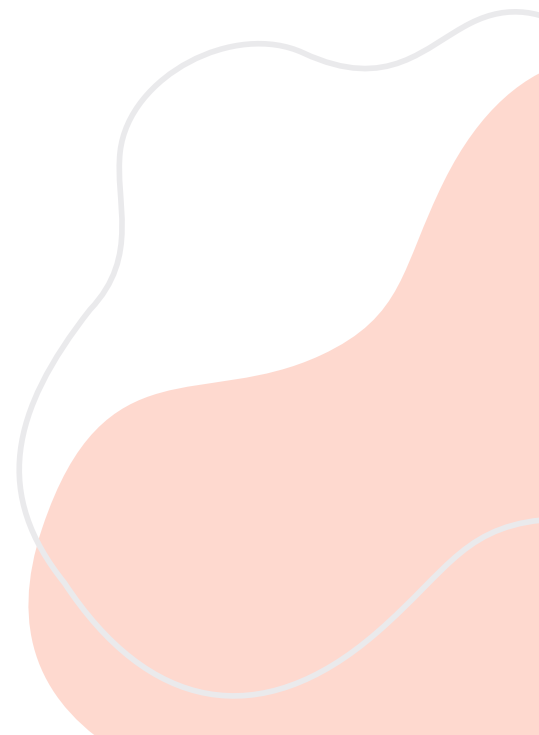
Business Objective:

- X Education seeks to identify the most promising leads by developing a predictive model.
 - The model will be designed to identify "hot leads," those with the highest likelihood of converting.
 - Once developed, the model will be deployed for future use in the lead generation and sales process.
 - Deployment of the model will enable X Education to consistently prioritize and focus on engaging with the most valuable leads.
 - This proactive approach will optimize their sales efforts and improve overall lead conversion rates.
- 
- 



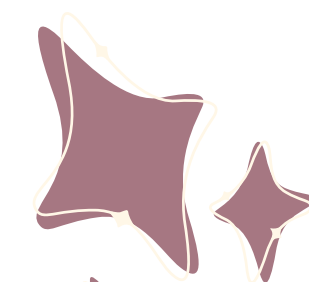
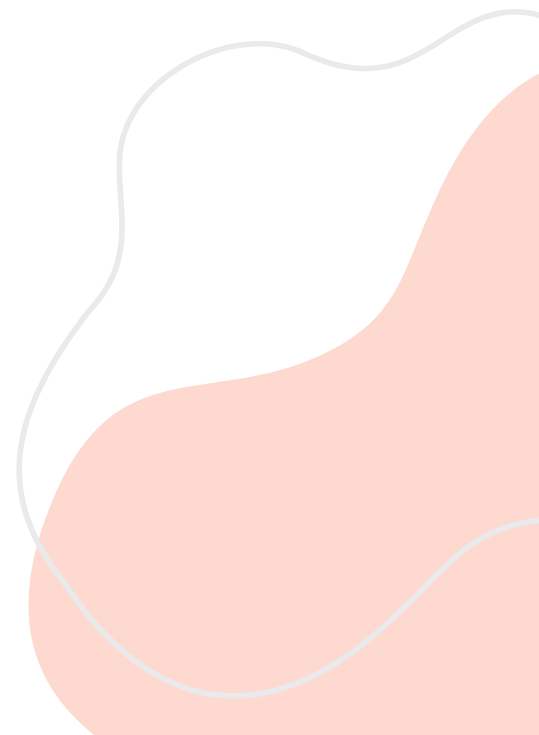
Strategy:



- Data Cleaning:
 - Handle duplicates, NA values, and missing data.
 - Drop irrelevant columns with many missing values.
 - Impute missing values and handle outliers.
 - Exploratory Data Analysis (EDA):
 - Analyze value counts, distributions, and correlations.
 - Feature Scaling & Encoding:
 - Scale features and encode categorical variables.
 - Classification Technique:
 - Use logistic regression for modeling and prediction.
 - Model Validation:
 - Validate model accuracy and reliability.
 - Presentation:
 - Present findings clearly and concisely.
 - Conclusions & Recommendations:
 - Draw insights and offer actionable recommendations.
- 



Data Manipulation

- Total Rows: 37, Total Columns: 9240.
 - Single value features such as "Magazine", "Receive More Updates About Our Courses", "Update me on Supply" have been removed.
 - Features like "Chain Content", "Get updates on DM Content", "I agree to pay the amount through cheque" have been dropped.
 - "Prospect ID" and "Lead Number" have been removed as they are unnecessary for analysis.
 - Features with low variance in value counts, such as "Do Not Call", "What matters most to you in choosing course", have been dropped.
 - Columns with over 35% missing values, like "How did you hear about X Education" and "Lead Profile", have been dropped.
- 
- 

Exploratory Data Analysis (EDA):

Total Rows: 37, Total Columns: 9240.

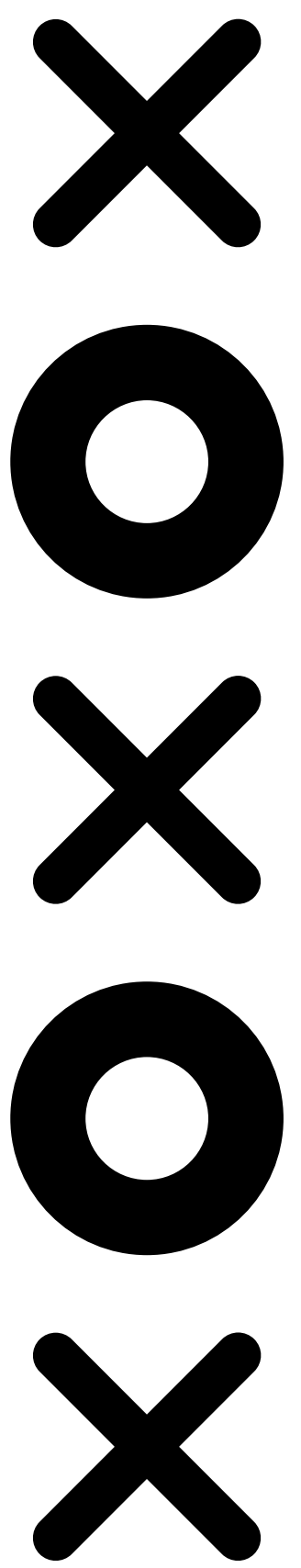
Single value features such as "Magazine", "Receive More Updates About Our Courses", "Update me on Supply" have been removed.

Features like "Chain Content", "Get updates on DM Content", "I agree to pay the amount through cheque" have been dropped.

"Prospect ID" and "Lead Number" have been removed as they are unnecessary for analysis.

Features with low variance in value counts, such as "Do Not Call", "What matters most to you in choosing course", have been dropped.

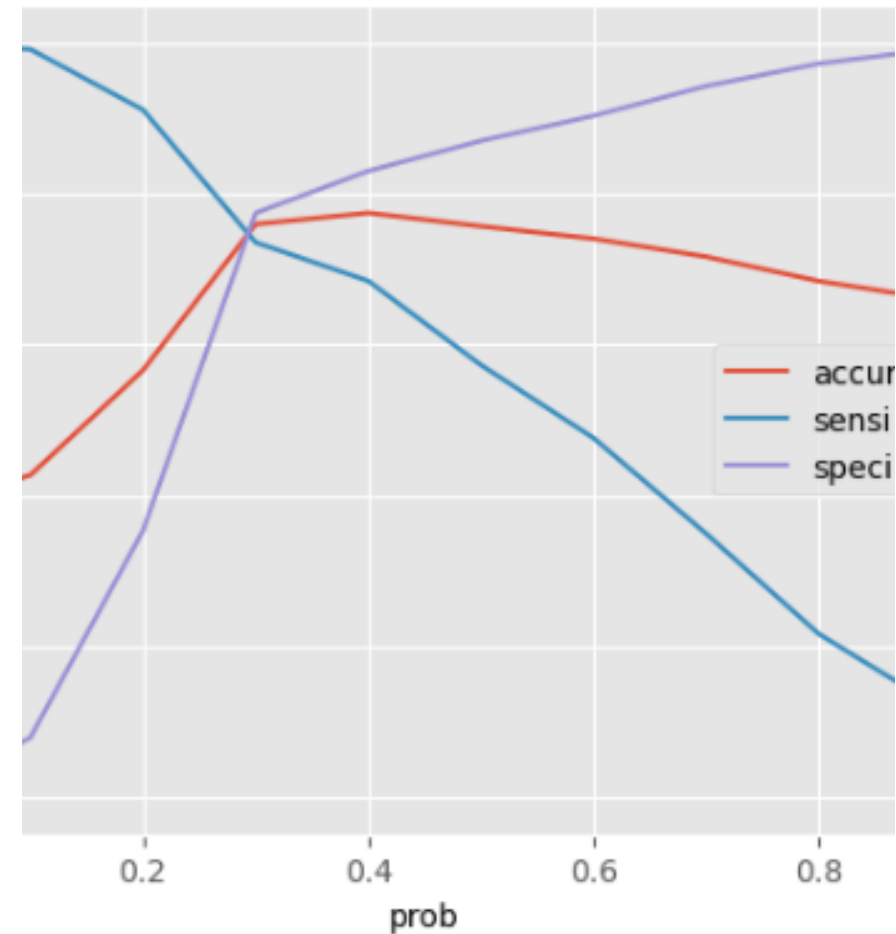
Columns with over 35% missing values, like "How did you hear about X Education" and "Lead Profile", have been dropped.



MODEL BUILDING

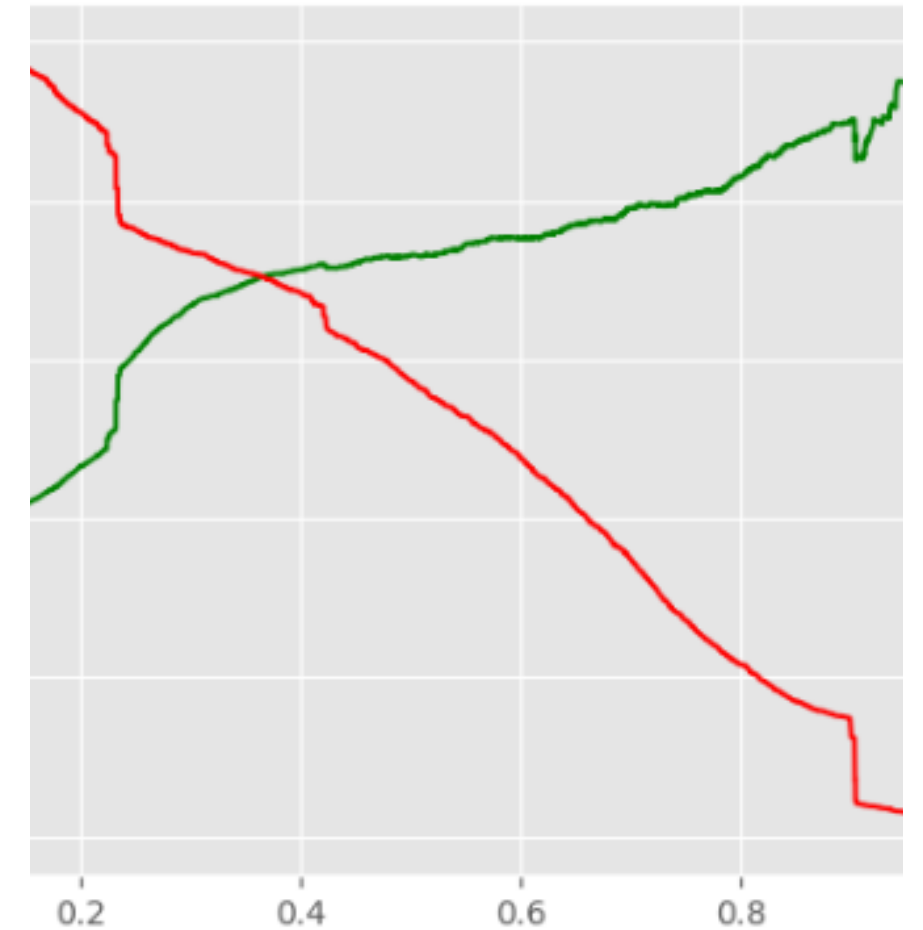
- Split the dataset into training and testing sets.
- Scale variables in the training set.
- Build the initial model.
- Use Recursive Feature Elimination (RFE) to eliminate less relevant variables.
- Build the next model with the selected variables.
- Eliminate variables based on high p-values.
- Check Variance Inflation Factor (VIF) for all remaining columns.
- Make predictions using the training set.
- Evaluate accuracy and other metrics.
- Predict using the test set.
- Perform precision and recall analysis on test predictions

Model Evaluation (Train)



Accuracy Sensitivity and Specificity

- Accuracy – 77.43%
- Sensitivity – 66.71%
- Specificity – 84.03%



Precision and Recall

- Precision – 71.72%
- ReCall – 62.73%

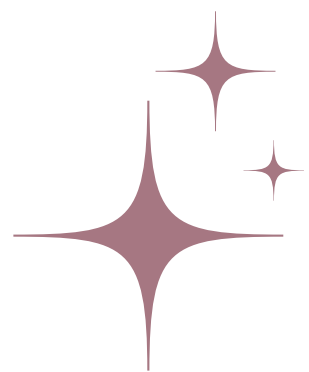
Conclusion

EDA:

- Leads spending more time than average are promising, suggesting targeted approaches for higher conversions.
- SMS messages and landing page submissions have a significant impact on lead conversion rates.
- Specializations in marketing and human resources show high conversion rates, indicating potential promising leads.
- Offering references or incentives for referrals can increase conversion rates.
- Alerts and information messages are associated with higher lead conversion rates.

Logistic Regression Model:

- The model demonstrates a high accuracy of approximately 77%.
- The threshold selection is based on accuracy, sensitivity, specificity, precision, and recall curves.
- Sensitivity and specificity are 66% and 84%, respectively.
- The model effectively identifies promising leads and those less likely to convert.
- Overall, the model proves to be accurate in predicting lead conversions.



**THANK
YOU!**

