X Education Lead Scoring Case Study

Project Done By:
Dhruv Shah
AC Janasivaguru
Sai Shruthi

X Education Company:

- An education company named X Education sells online courses to industry professionals.
- On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.
- When these people fill up a form providing their email address or phone number, they are classified to be a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.

Problem Statement & Objective of the Study

Problem Statement:

- X Education gets a lot of leads, its lead conversion rate is very poor at around 30%
- X Education wants to make lead conversion process more efficient by identifying the most potential leads, also known as Hot Leads
- Their sales team want to know these potential set of leads, which they will be focusing more on communicating rather than making calls to everyone.

Objective of the Study:

- To help X Education select the most promising leads, i.e., the leads that are most likely to convert into paying customers.
- The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO has given a ballpark of the target lead conversion rate to be around 80%.

Table of Contents:

"Optimizing Lead Conversion with Advanced Lead Score Analysis"

Introduction

- Introduction to the challenge faced by X Education – high lead volume but low conversion rate.
- Importance of efficient lead scoring for improving conversion rates.
- Brief mention of the solution provided by building a predictive lead scoring model.

Business Goals

- Detailed explanation of the business goal: To identify and prioritize 'Hot Leads' for increased conversion rates.
- The potential impact on revenue and customer acquisition through targeted communication.
- Alignment of business goals with the implementation of the lead scoring model.

Lead Scoring Process

- Overview of how leads are generated via various channels like websites, referrals, and past referrals.
- Explanation of the lead scoring process based on user behavior, such as form fills, course browsing, and video views.
- Explanation of lead scoring based on user interactions such as form fills, course browsing, and video views.
- Visualization of the lead scoring process to highlight its importance in the sales funnel.

Data Overview

- Highlights from the dataset, emphasizing key features used for lead scoring.
- Challenges faced with the existing lead conversion rate and the need for a predictive model to address it.
- The rationale for adopting a predictive model to address these challenges.

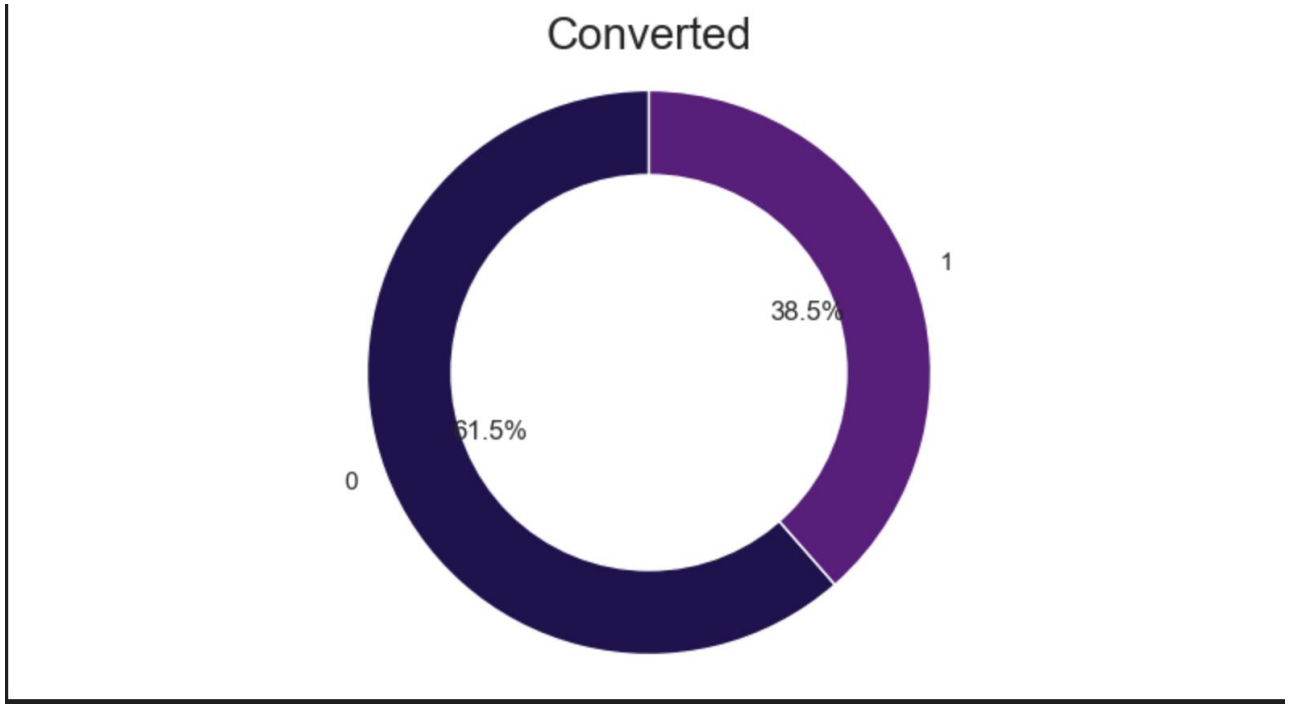
Data Cleaning and Feature Engineering

- Explanation of the data cleaning steps, including dropping unnecessary columns and handling asymmetrical indices.
- Transformation of binary categorical variables to numeric form.
- Emphasis on how clean data improves model performance.

Exploratory Data Analysis (EDA)

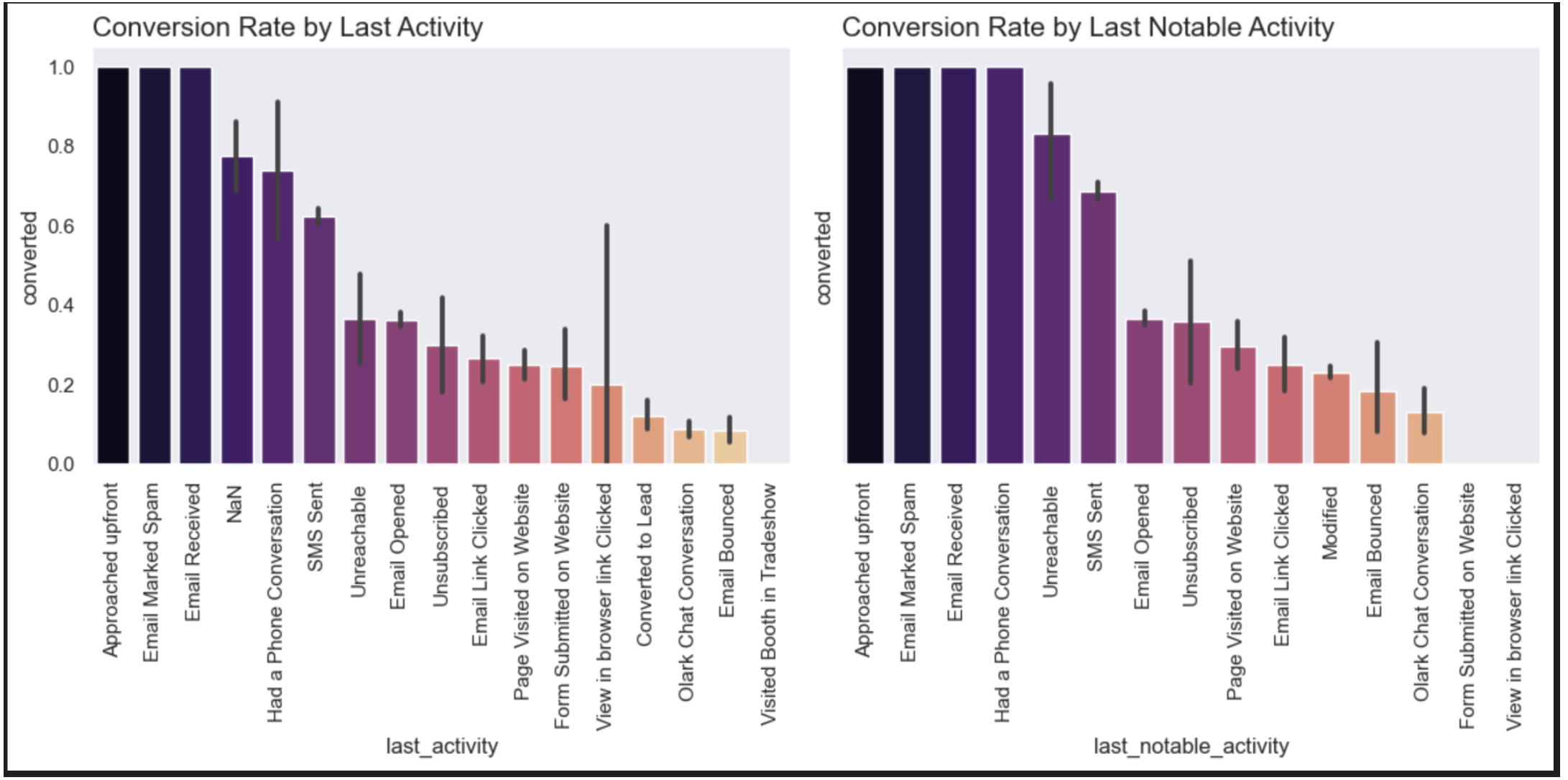
- Visualizations showcasing insights from EDA, including distribution of lead sources, last activities, and specializations.
- Impact analysis of missing values on conversion rates.
- Interactive visualizations to engage the audience with key findings.
- Throughout our EDA, we identified key features such as phone interactions, email interactions, and profile score, which were the most important features in determining the conversion rate of a lead.

Exploratory Data Analysis



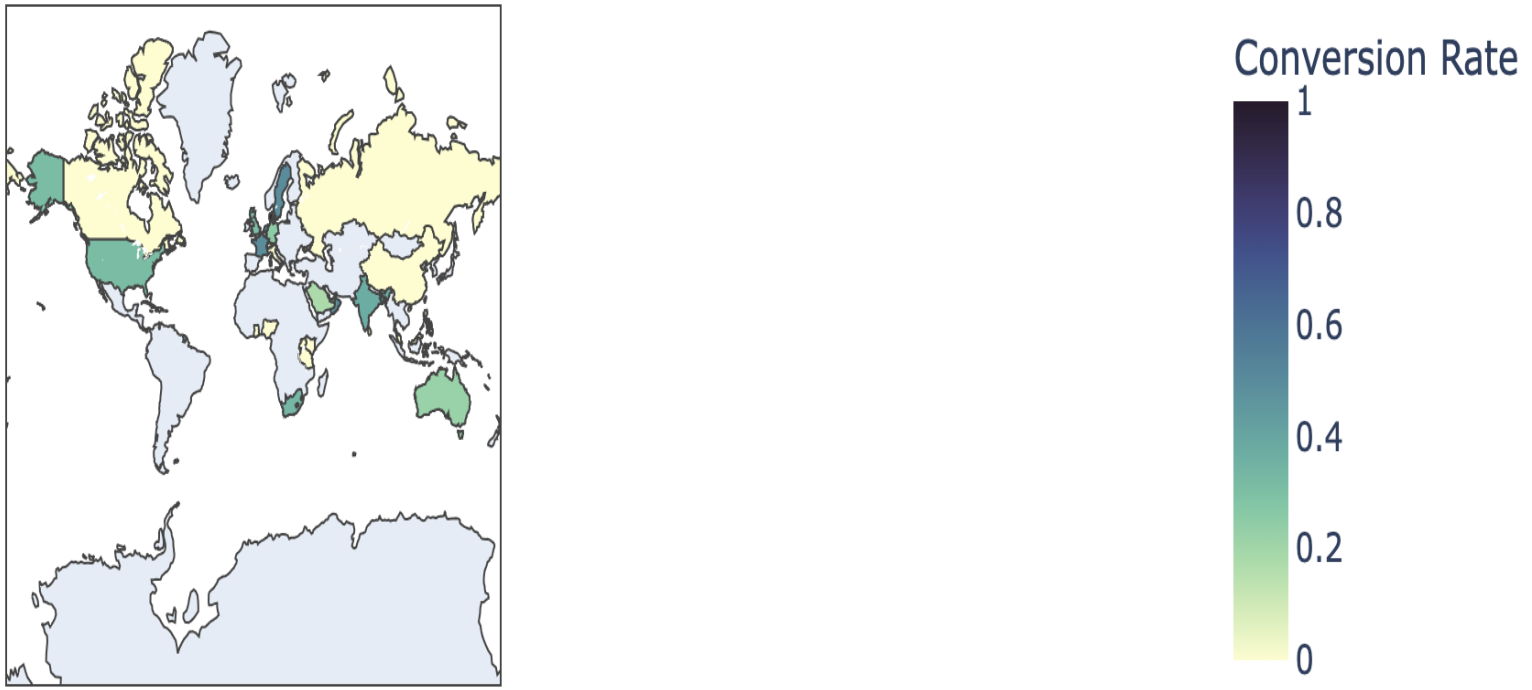
	asymmetrique_activity_index	asymmetrique_profile_index	asymmetrique_activity_score	asymmetrique_profile_score
asymmetrique_activity_index	1.000000	-0.145399	0.855985	-0.122669
asymmetrique_profile_index	-0.145399	1.000000	-0.145366	0.883177
asymmetrique_activity_score	0.855985	-0.145366	1.000000	-0.114636
asymmetrique_profile_score	-0.122669	0.883177	-0.114636	1.000000

Exploratory Data Analysis

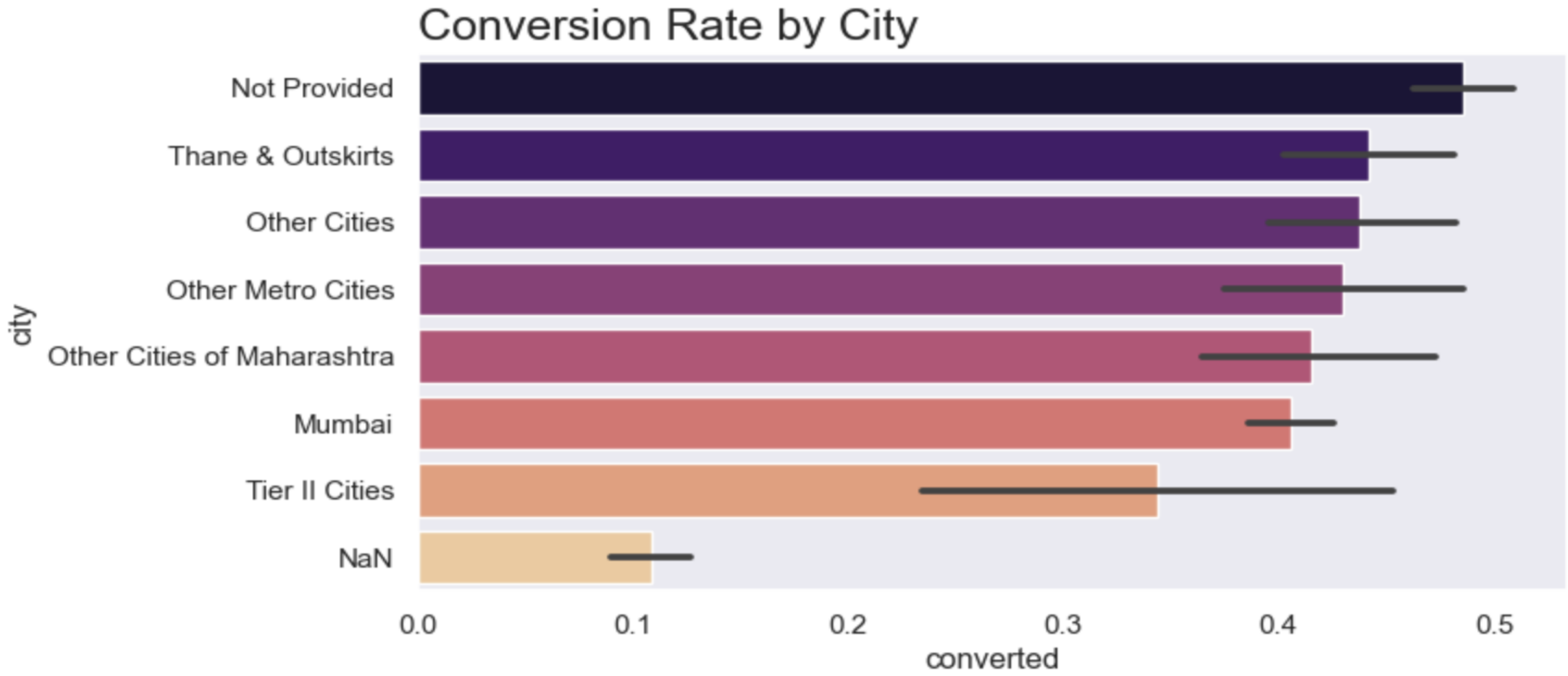


Exploratory Data Analysis

Conversion Rate by Country



Exploratory Data Analysis



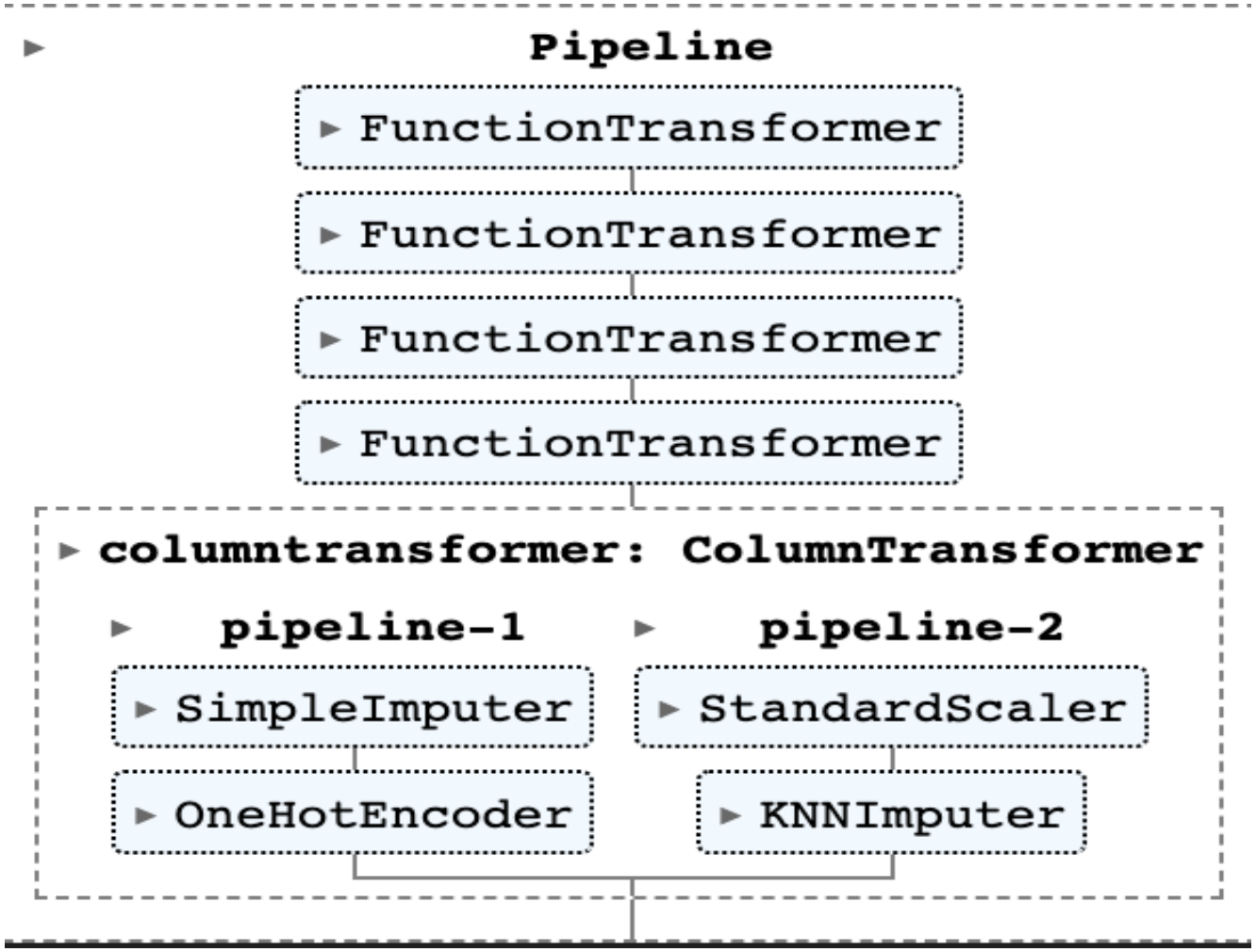
Outlier Handling

- Explanation of the approach taken to handle outliers in numerical features, emphasizing their impact on model performance.
- Discussion on the impact of outliers on model performance and the importance of their removal.

Preprocessing and Model Building

- Overview of preprocessing steps, including one-hot encoding and standard scaling.
- Introduction to logistic regression as the chosen model for lead scoring.

Data Wrangling



Model Evaluation

- Cross-validation scores and performance metrics for logistic regression.
- Comparison with other models (e.g., Random Forest, Decision Tree) and the rationale behind choosing logistic regression.
- Explanation of why these metrics are essential for evaluating model effectiveness.

Model Validation and Testing

- Application of the model to the test set.
- Evaluation metrics on the test set for logistic regression and Random Forest.
- Confidence in model performance through testing results.

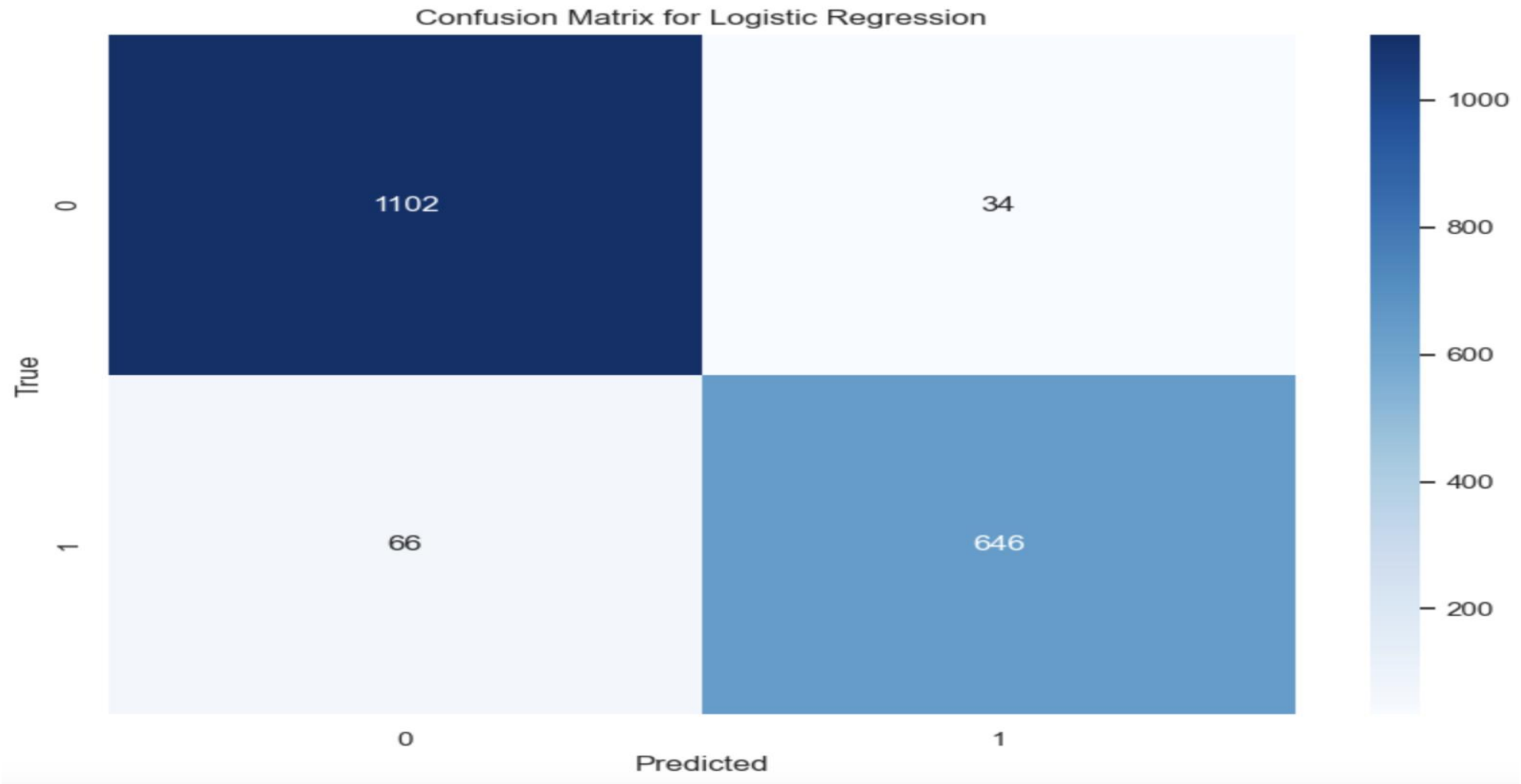
Confusion Matrices

- Visual representation of confusion matrices for both Logistic Regression and Random Forest models.
- In-depth discussion on true positives, true negatives, false positives, and false negatives.
- Interpretation of confusion matrices to assess model behavior.
- In Depth analysis of X education company provides the company with actionable insights to work on and improve their lead conversion rate.

Results and Recommendations

- Presentation of lead scoring results, including predicted probabilities.
- Recommendations on how to leverage the model to prioritize 'Hot Leads' for the sales team.
- Practical insights derived from the results.
- As is evident by the results, our model is performing well above the 80% threshold set by the company, in fact it is over 90% accurate, which is a great result.

Comparison of Models



Business Impact

- Discussion on the potential positive impact of the lead scoring model on business outcomes.
- Projected increase in conversion rates, revenue, and the overall efficiency of the sales process.
- Alignment of business impact with initial goals.
- The cherry on top of our data science project is that we have now automated the lead scoring process by implementing a pipeline and a model that can be used to score leads in the future.

Conclusion

- Summary of key findings, outcomes, and the successful implementation of the lead scoring solution.
- Encouragement for continuous optimization and improvement.
- This will no doubt be an immense help to the sales team, as they will now be able to focus on the most promising leads, resulting in a higher conversion rate and more revenue for the company.


THANK YOU