Analysis Of
Lead Score Case Study

Presented by

Nasrin Khatun nasrinofficial786@gmail.com
Niranjan Muralidhar niranjanmuralidhar@gmail.com
Zubair Mohammed mohdzubair34@gmail.com

Problem Statement

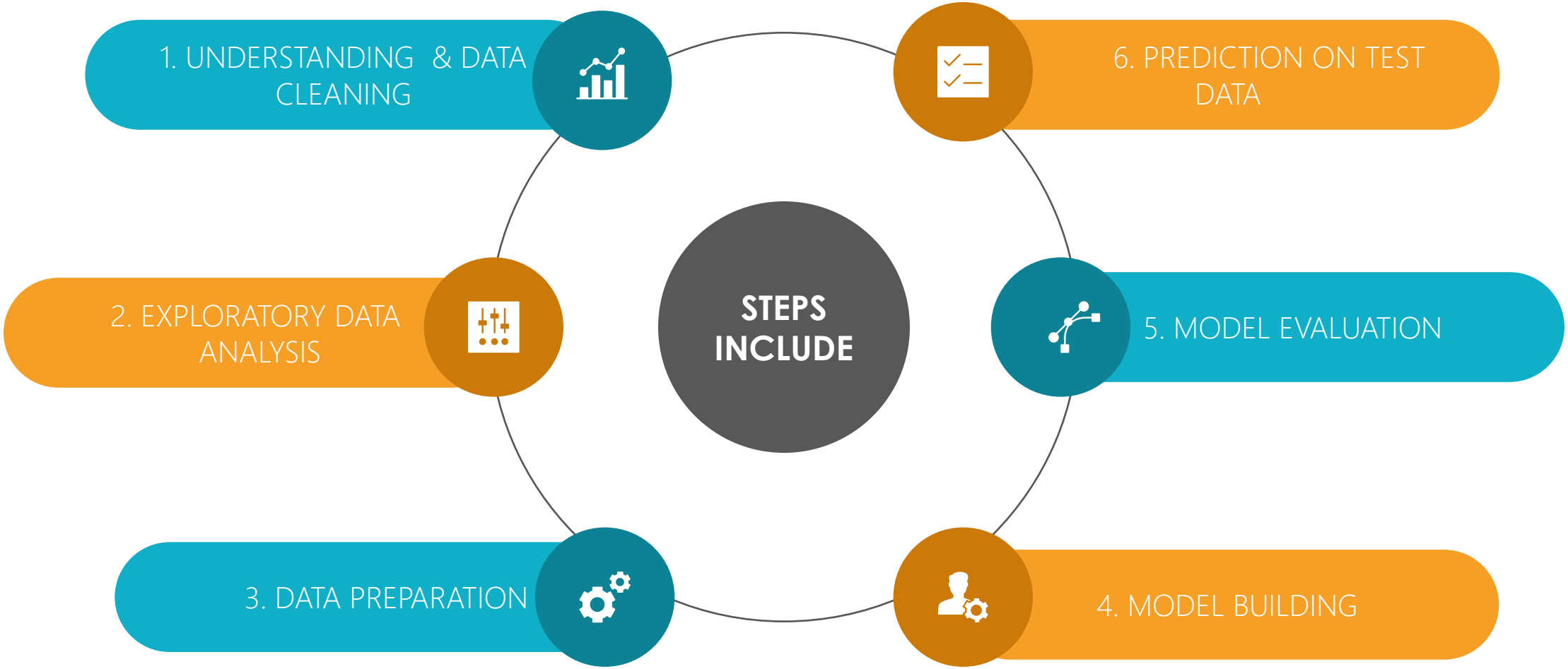
X Education offers online courses tailored for industry professionals. The organization promotes its offerings across various websites and search engines, including Google. Upon visiting the website, potential customers may explore the available courses, complete a registration form, or view promotional videos. When individuals submit a form with their email address or phone number, they are categorized as leads. Additionally, the company generates leads through referrals from previous clients.

Once leads are obtained, members of the sales team initiate contact through phone calls and emails. This engagement process results in the conversion of some leads, although a significant portion remains unconverted. The average lead conversion rate for X Education is approximately 30%.

Business Goal

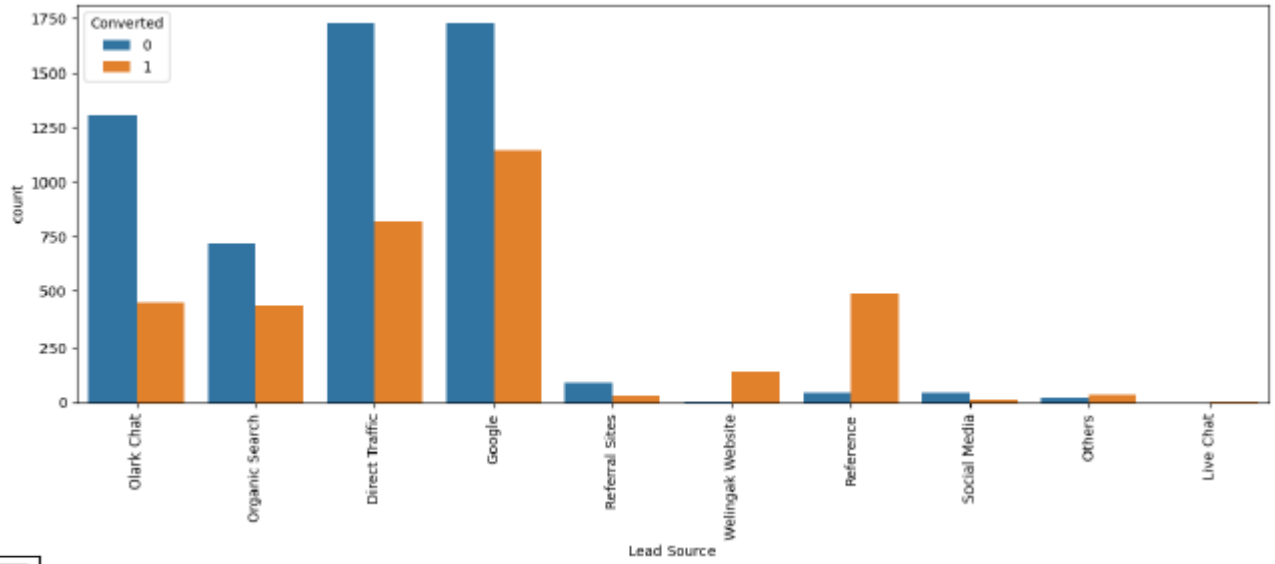
X Education requires assistance in identifying the most promising leads, specifically those with the highest likelihood of converting into paying customers. The organization seeks to implement a model that assigns a lead score to each potential customer, ensuring that those with elevated lead scores exhibit a greater probability of conversion, while those with diminished lead scores demonstrate a lower probability. The CEO has indicated a target lead conversion rate of approximately 80%.

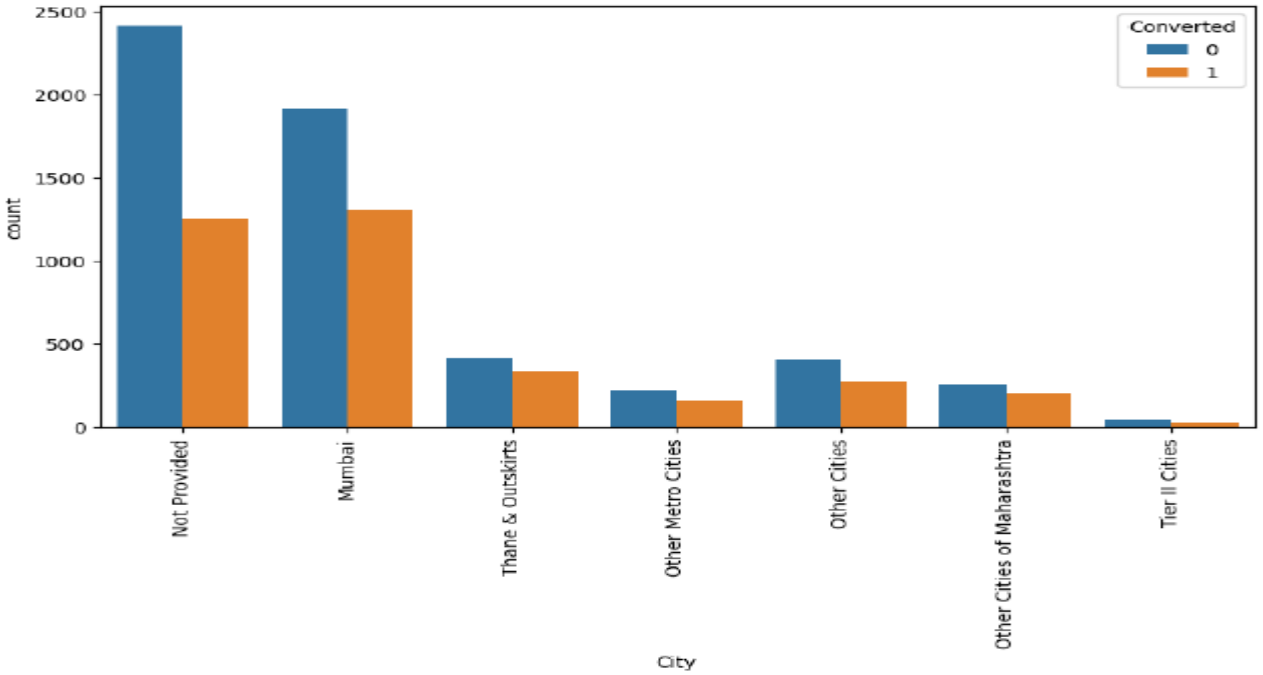
Case Stusy Analysis



Exploratory Data Analysis

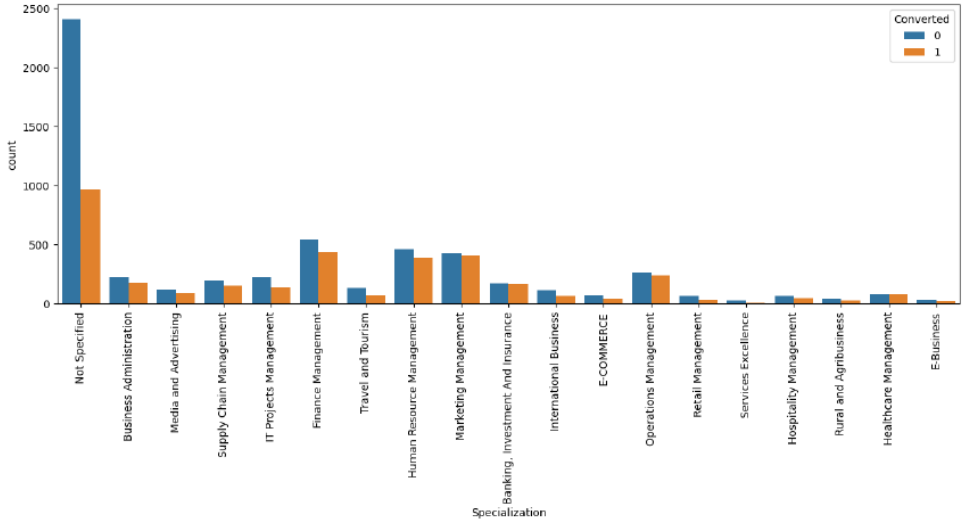
Lead sources such as ‘Reference’ and ‘WelingakWebsite’ show exceptionally high conversion rates.The majority of leads originate from ‘Direct Traffic’ and ‘Google’.





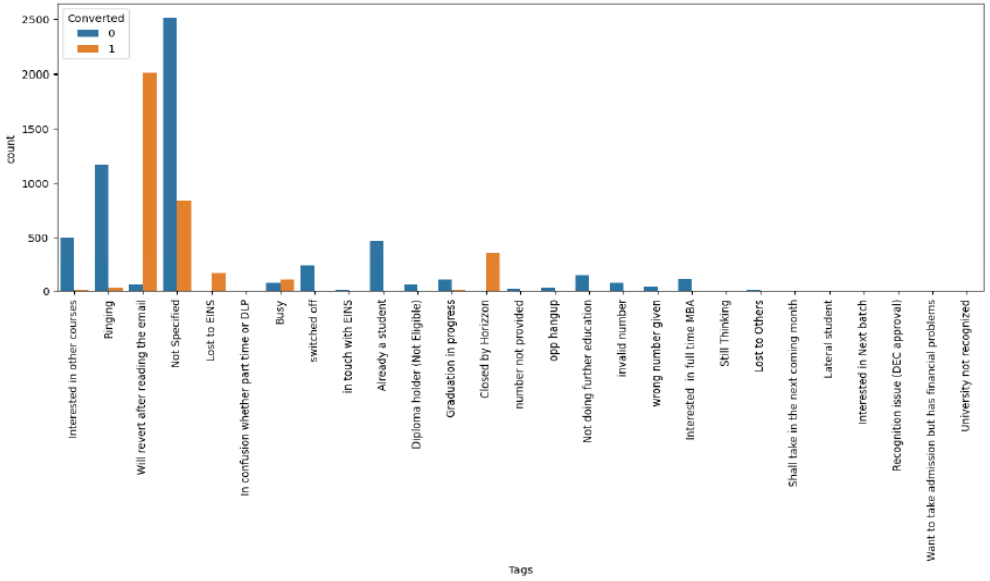
The majority of leads are from Mumbai or have missing values (Not Provided). - Since this column does not provide significant insights for our analysis, we have decided to drop it.

Exploratory Data Analysis

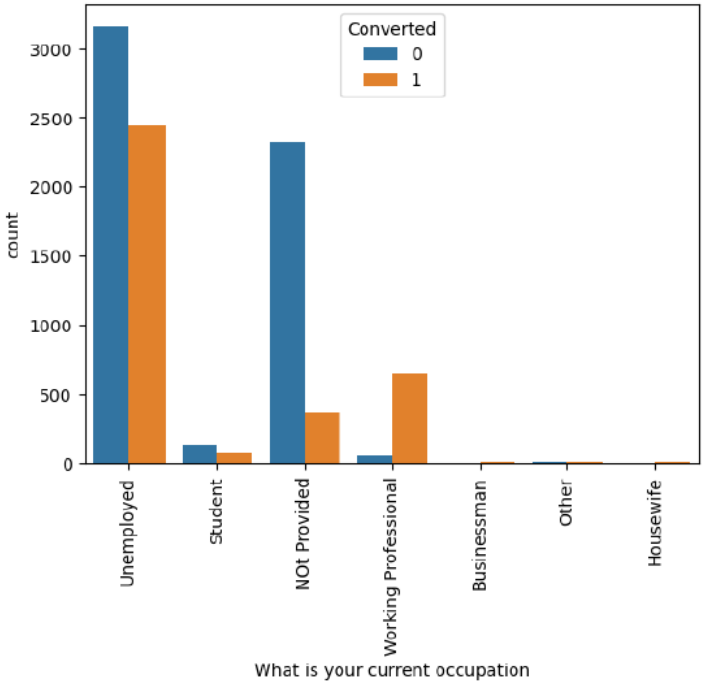


Leads from Management Specialization have a higher probability of conversion. - This indicates a strong preference among prospective students, making it a key factor in lead prioritization.

The following tags show high conversion rates: "Will revert after reading the email," "Not Specified," "Closed by Horizon," "Lost to EINS," and "Busy."

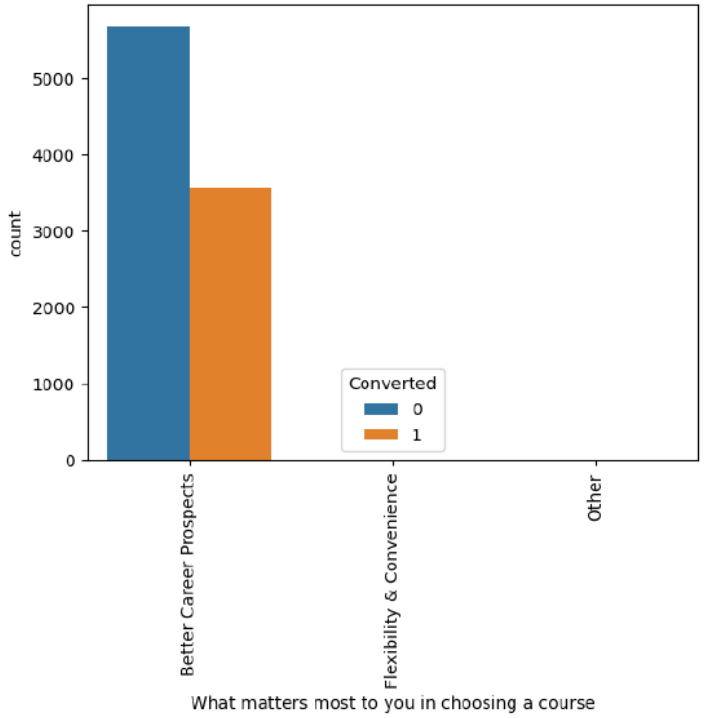


Exploratory Data Analysis

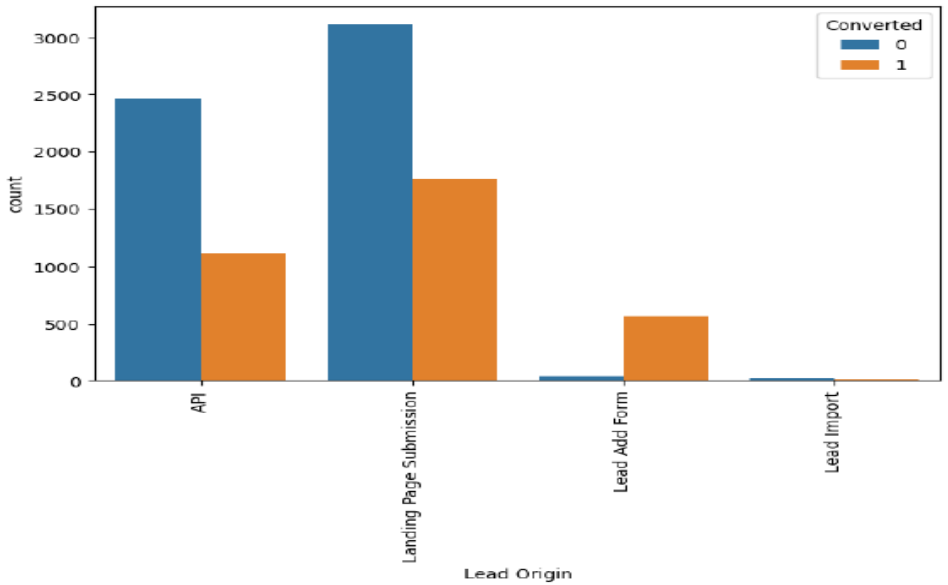


Current Occupation: Working professionals pursuing the course have a higher likelihood of enrollment. Unemployed leads make up the majority in absolute numbers.

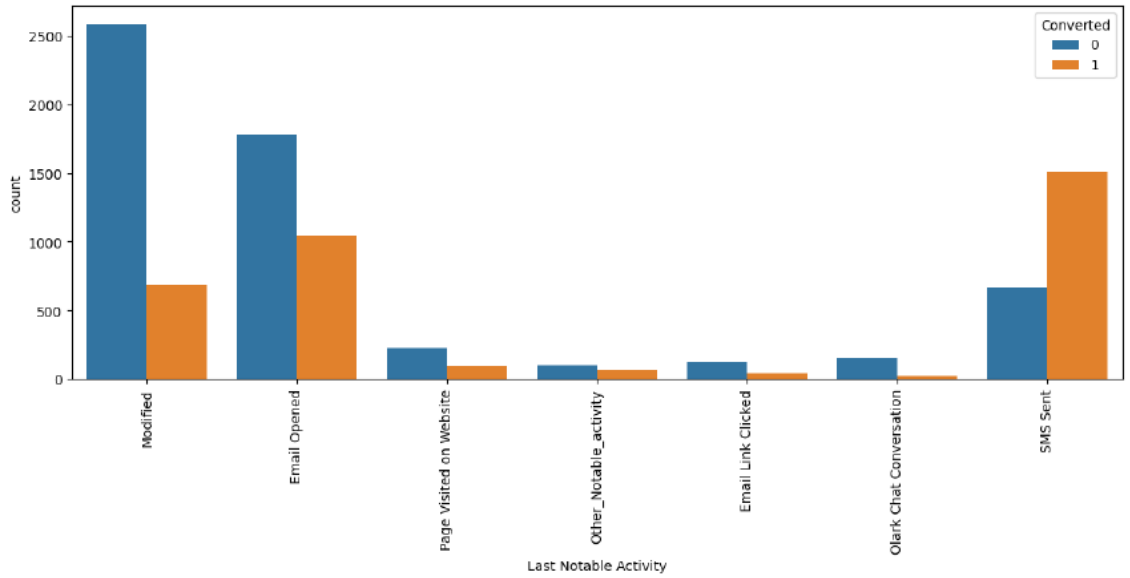
Key Factors in Choosing a Course: Better Career Prospects is the most significant factor influencing enrollment.Flexibility & Convenience are negligible, and other categories hold no value, making it suitable to append or drop.



Exploratory Data Analysis

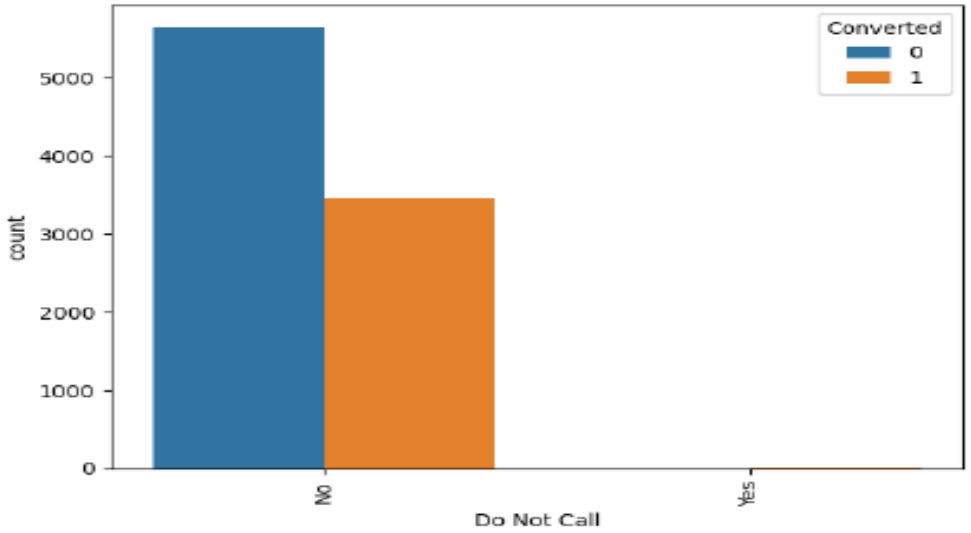


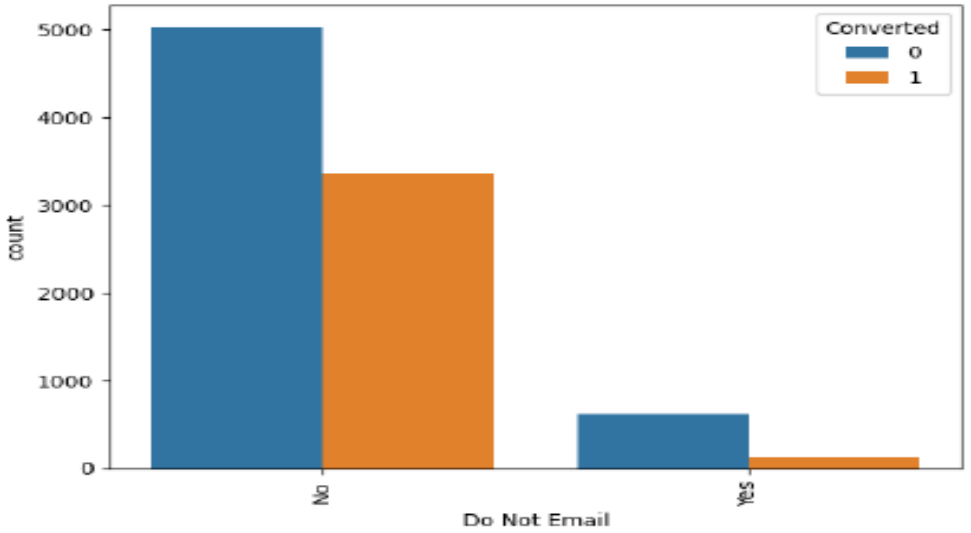
API and Landing Page Submission generate the highest number of leads and conversions, while Lead Add Form has a high conversion rate but a lower lead count.



Last Notable Activity:'SMS Sent' has the highest conversion rate, followed by 'Email Opened' and 'Modified

Exploratory Data Analysis

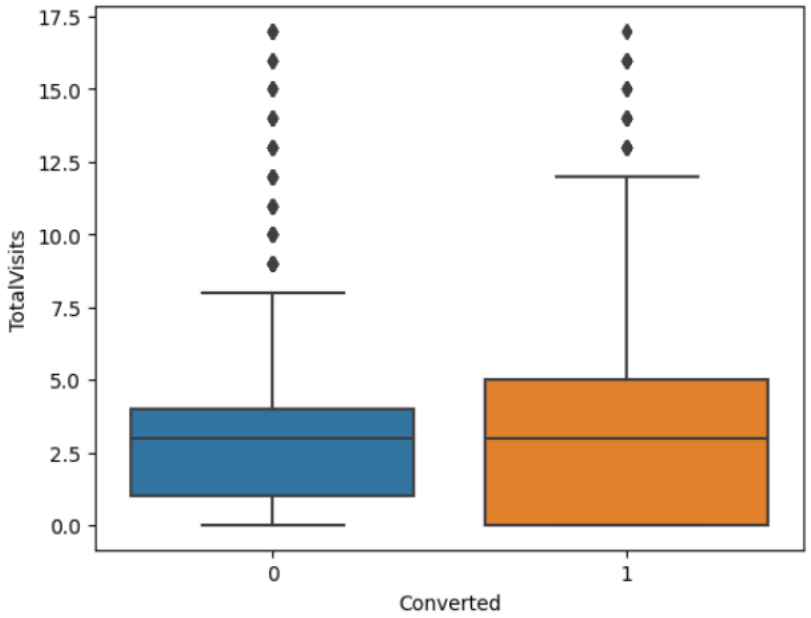


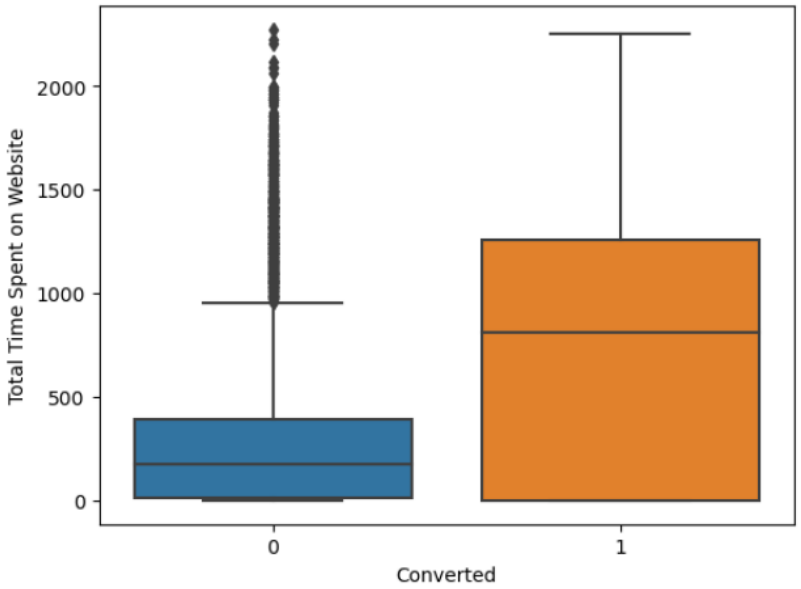


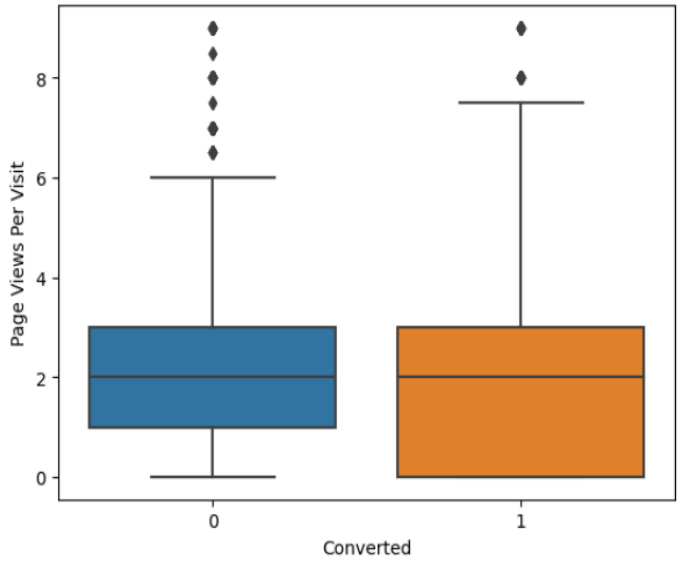
Do Not Call: This column can be dropped as more than 90% of its values are the same, providing little to no predictive value.

Do Not Email: This column also can be dropped as its skewing graph shows highly imbalanced data.

Exploratory Data Analysis







Out of 'Total visits', 'Total Time spent on Website', 'Page views per visit'. People Spending more time on website are more likely to get converted.

Project Analysis

DATA PREPARATION

1. Handling missing values
2. Encoding categorical variables
3. Handling Outliers
4. Creating dummy variables for categorical features

FEATURE SCALING

1. Standardization() : Used StandardScaler() to normalize numerical variables.

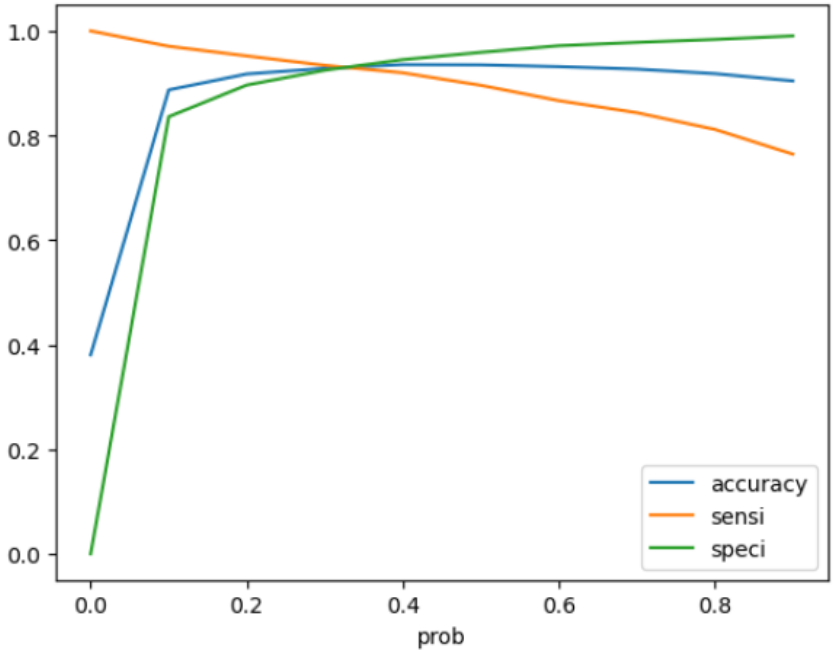
SPLITTING DATASET

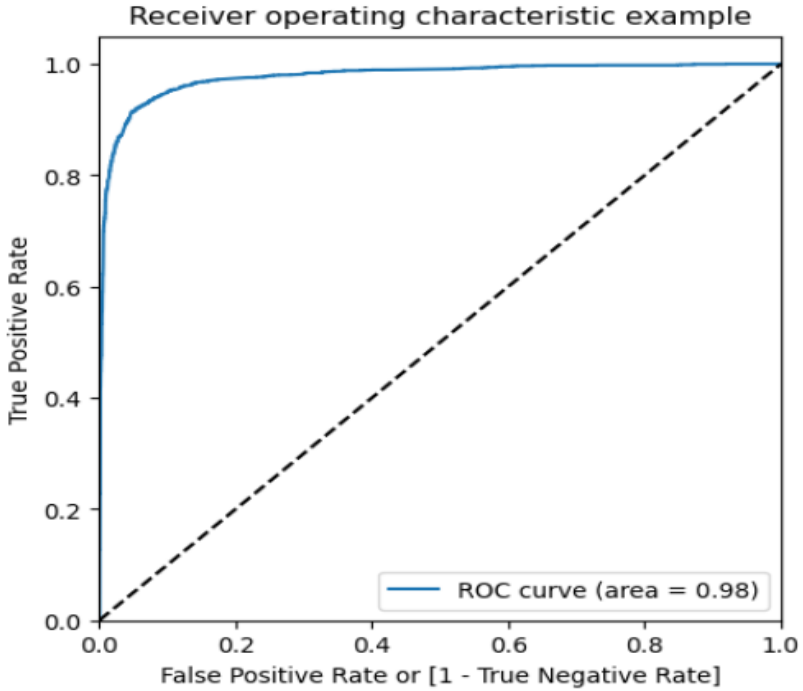
1. Used train_test_split to split the dataset into train dataset and test dataset with 70-30 ratio for fair evaluation.

MODEL BUILDING

1. Feature selection using RFE.
2. Logistic Regression for predicting lead conversion.
3. Adjusted the decision threshold to balance the lead conversion strategy.

MODEL EVALUATION
(TRAIN DATA)





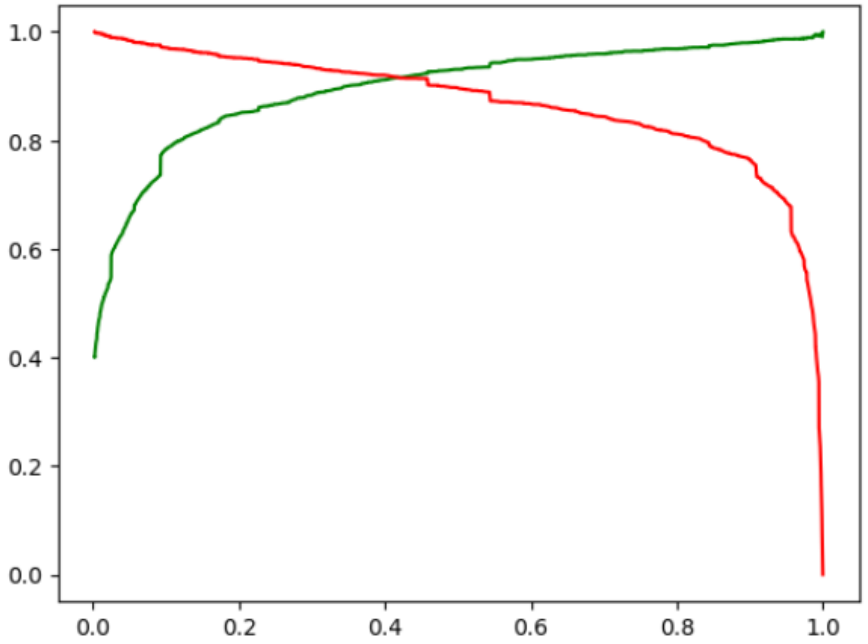
This evaluation was carried out using the Confusion Matrix on three key metrics:
1.Accuracy : 93.50%
2.Sensitivity : 89.60%
3.Specificity : 95.90%
We chose an arbitrary cutoff of 0.5 to determine the predicted values, based on the probabilities provided by the model. If the probability was less than or equal to 0.5, the predicted value would be 0; otherwise, it would be 1.

We used the ROC Curve and the specificity-sensitivity trade-off to determine the optimal cut-off point of 0.3 for our model, with the ROC Curve yielding an area of 0.98, indicating the model's efficiency

With confusion Matrix :

3594	288
154	2231

MODEL EVALUATION
(PREDICTION)



We applied the cut-off point f 0.3 to make predictions on both our training and test sets and then compared the model's performance for each.

Property	Training Data	Testing Data
Precision	88.56%	89.49%
Recall	93.54%	94.45%

Accuracy	93.74%
Sensititvity	94.45%
Specificity	93.31%

With confusion Matrix :

1564	112
56	954

OUTCOME

- This classification model can be deployed to help the business predict outcomes, such as lead conversion, with 96% accuracy.
- High precision and recall for both classes ensure the model minimizes false positives (incorrectly predicting a lead will convert) and false negatives (failing to identify actual lead conversions).
- A strong ROC AUC score shows the model's ability to effectively differentiate between both classes, further enhancing its performance.

CONCLUSION

It is found that the variables that mattered the most in the potential buyers are (In descending order) :

- Lead Source_Google
- Total Time Spent on Website
- Tags_Closed by Horizzon
- Lead Source_Welingak Website
- What is your current occupation_NOt Provided
- Tags_Lost to EINS
- Tags_Ringing

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

Thank You