Lead Scoring Case Study

Mukul Kumar Samal
Aazeenelsa

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Business Objective:

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

Solution Methodology

1. **Data Cleaning:** The provided dataset was cleaned by addressing missing values and irrelevant options in categorical variables.

2. **Exploratory Data Analysis:** EDA was performed to understand data distribution, identify outliers, and uncover patterns within the data.

3. **Feature Engineering:** Dummy variables were created for categorical features, and numerical features were standardized.

4. **Model Building:** A predictive model was built using a robust classification algorithm. Feature selection techniques like Recursive Feature Elimination and analysis of Variance Inflation Factor were employed to select the most relevant predictors.

5. **Model Evaluation:** The model demonstrated excellent predictive performance on both training and test datasets:

   o **Accuracy:** 92.29%, 92.78%

   o **Sensitivity:** 91.70%, 91.98%

   o **Specificity:** 92.66%, 93.26%

   o **ROC AUC:** 0.97

6. **Threshold Optimization:** The optimal probability threshold for classification was determined to be 0.3, achieving a balance between precision (89%) and recall (91%) on the test data.
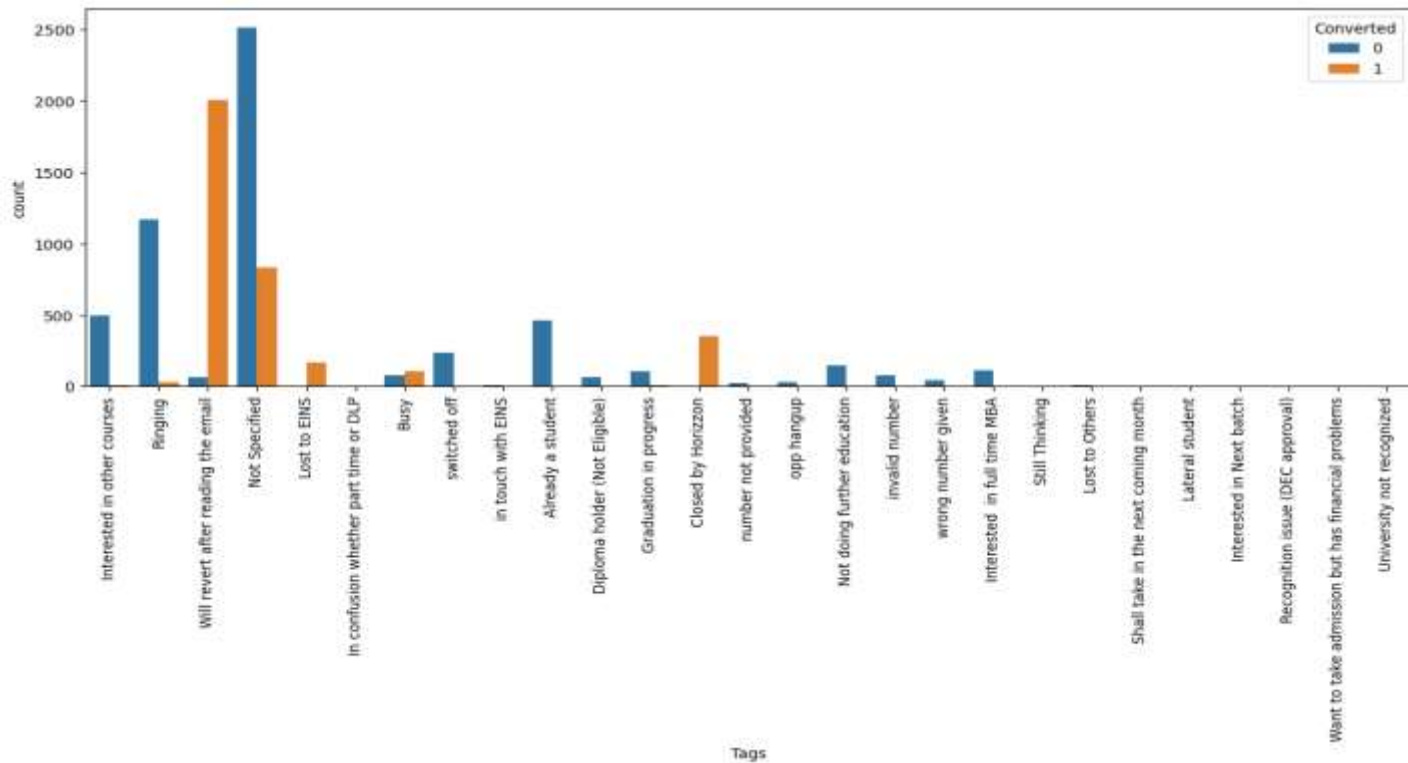
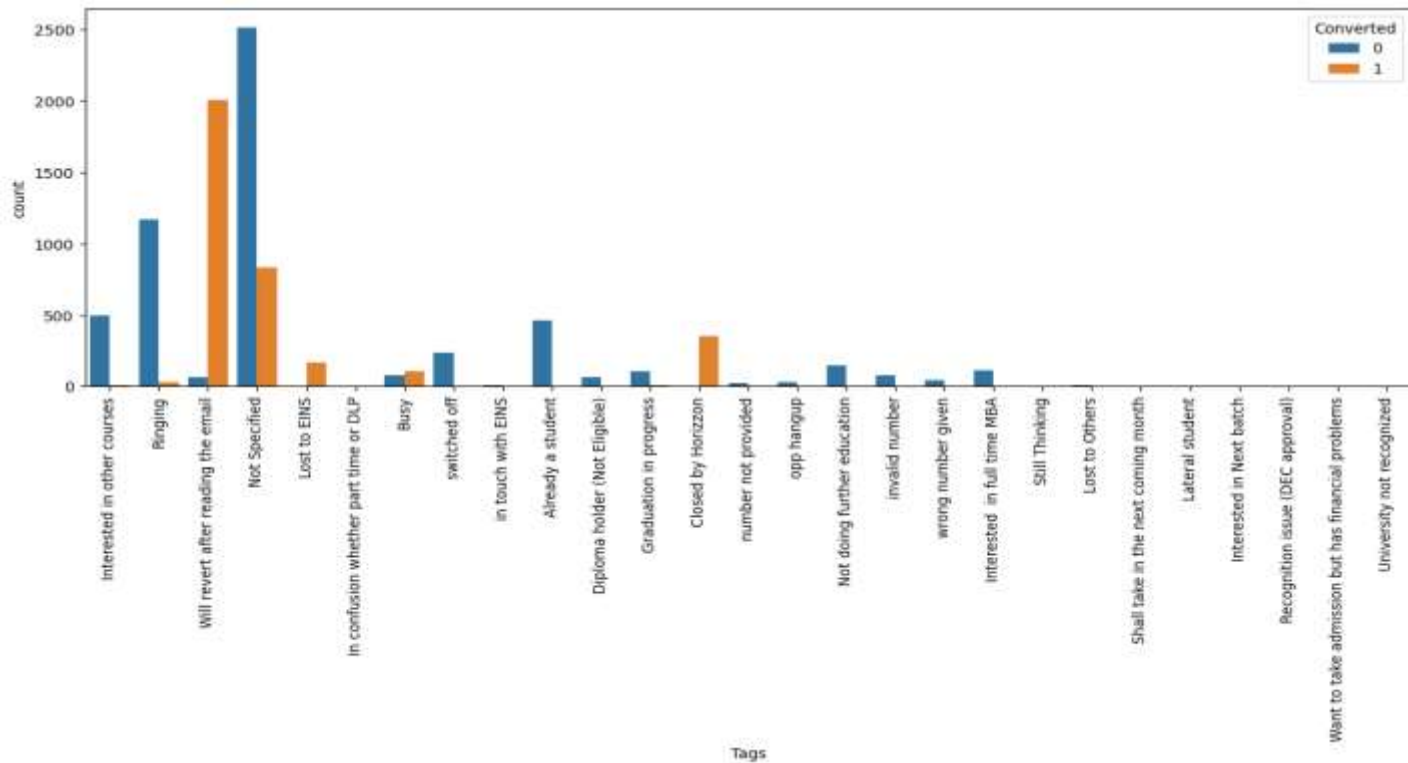# EDA - Last Notable Activity
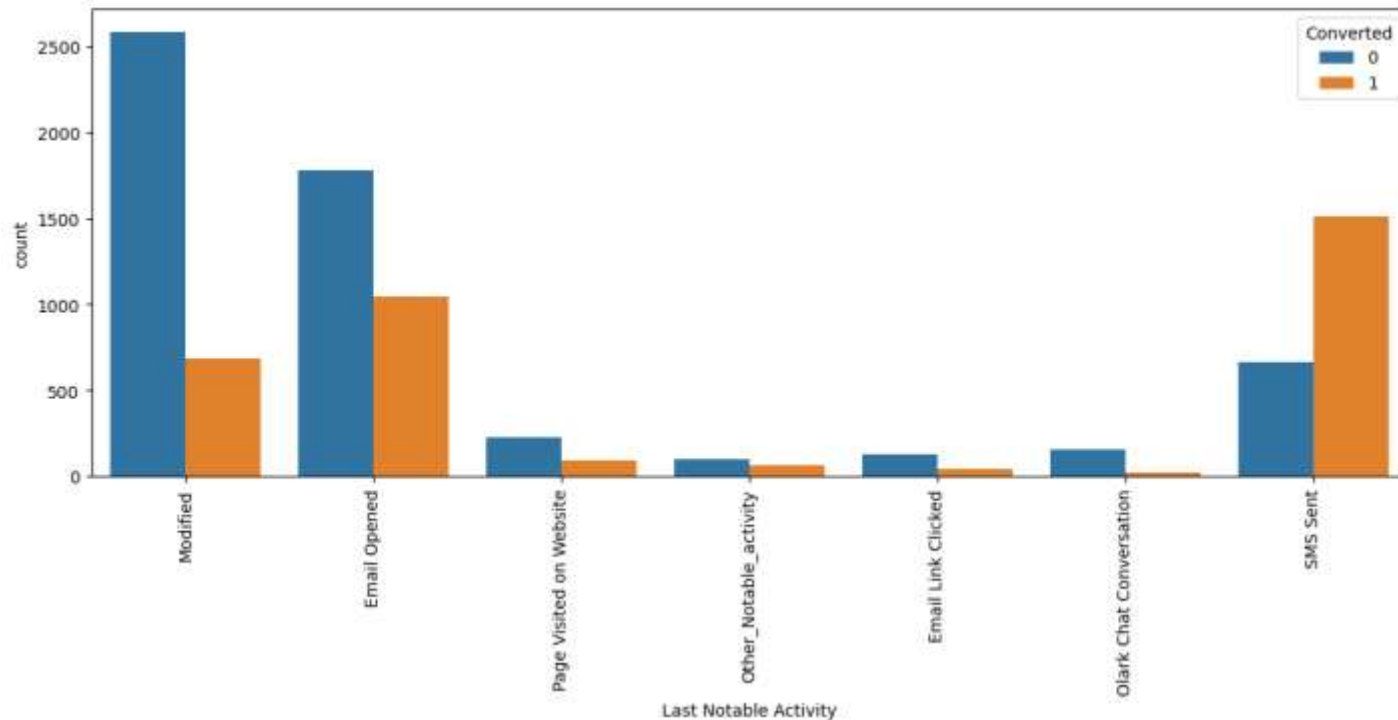
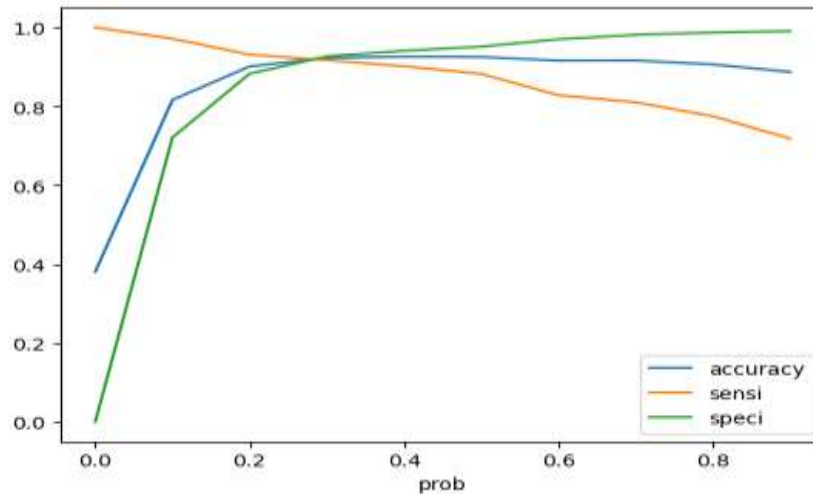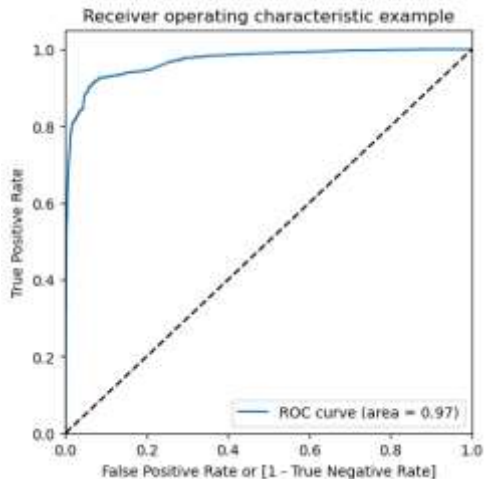# EDA - Correlation of converted values
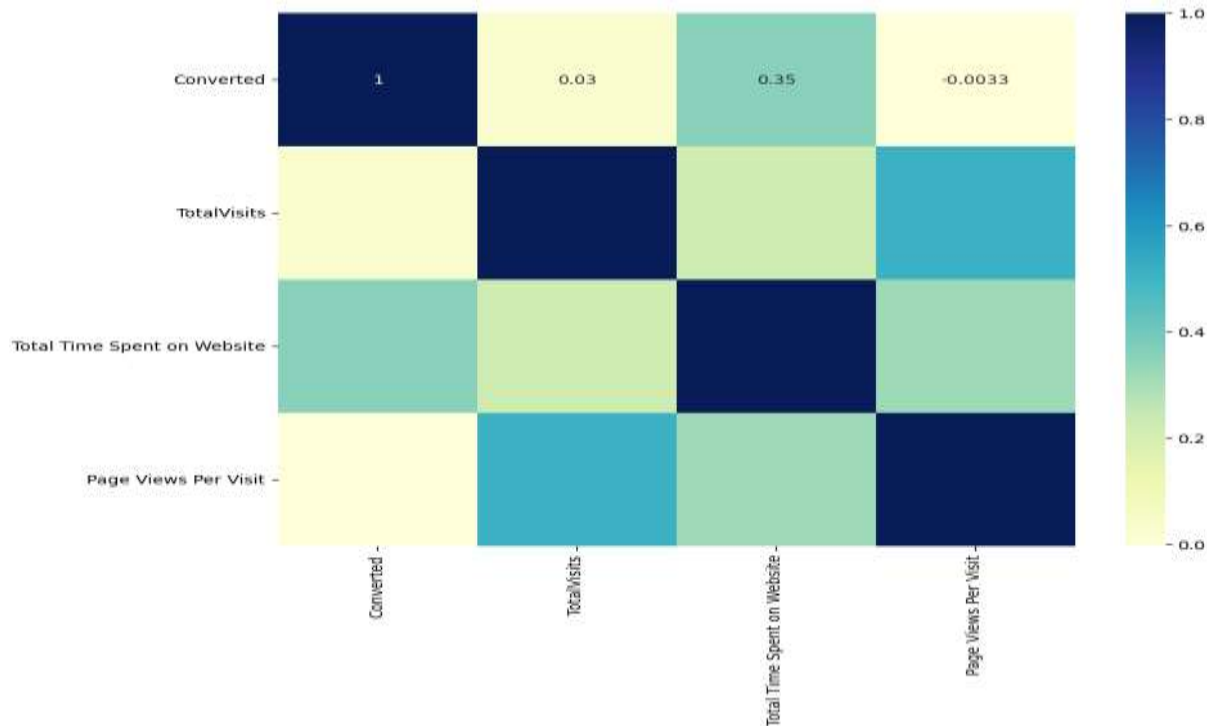
EDA - Categorical variables converted

EDA - Categorical variables converted

ROC Curve

Finding optimal cutoff point

- From the 2nd graph we understand that the optimal cutoff point is 0.3

Conclusion and Recommendations

- While the summary provides a robust overview of the methodology and model performance, it lacks specific insights into the key factors driving enrolment and actionable recommendations.

- **Translate insights into actionable strategies:** For example:

    – **Website Optimization:** If certain web pages or content were found to be highly engaging for potential enrollees, suggest improvements to those areas.

    – **Marketing and Outreach:** If specific referral sources or marketing channels proved most effective, recommend focusing efforts on those avenues.

    – **Course Design:** If certain course features or topics were highly appealing, suggest highlighting those aspects in marketing materials.

By incorporating these details, the report will provide the CEO of X Education with a clear understanding of the factors driving enrolment and actionable recommendations to attract more industry professionals to their courses.