Lead Scoring Case Study

Harshada Salvi & Arijit Das

Problem Statement

X Education, an online education company, faces a challenge with a low lead conversion rate (30%). To enhance efficiency, they aim to identify 'Hot Leads' among the numerous daily inquiries. 'Hot Leads' are individuals with a higher likelihood of conversion. These leads can be those who actively browse courses, fill out forms, or come through referrals. By focusing the sales team's efforts on these potential leads, X Education aims to increase its overall lead conversion rate, ultimately improving the effectiveness of their sales outreach.

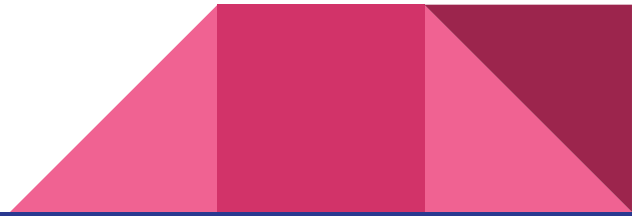
Business Objective

Lead X envisions a model that assigns lead scores (ranging from 0 to 100) to each inquiry, facilitating the identification of 'Hot Leads' for increased conversion rates. The CEO aims for an ambitious 80% lead conversion rate.

To address future constraints, such as peak times and resource optimization, the model should provide actionable insights. It should guide the company on deploying manpower efficiently during peak periods and recommend strategies to meet or exceed the 80% conversion target. Additionally, the model should offer post-achievement approaches, outlining how to sustain and further improve conversion rates once the target is met. The objective is to create a dynamic and adaptable solution that not only identifies potential leads but also guides strategic decision-making in response to varying operational challenges.

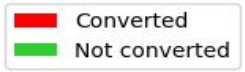
Problem Approach

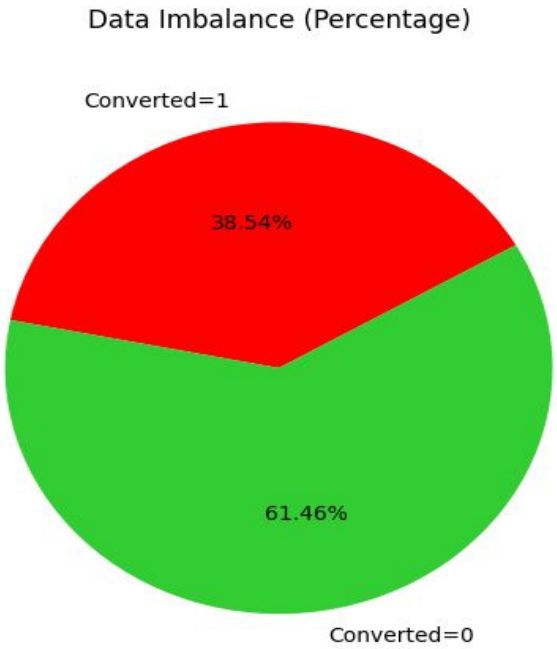
1. Cleaning of the data
2. EDA-Exploratory Data Analysis
3. Creating Dummy Variables
4. Test-Train Split
5. Model Building
6. Model Evaluation
7. Making Predictions on the Test Set



Implementation

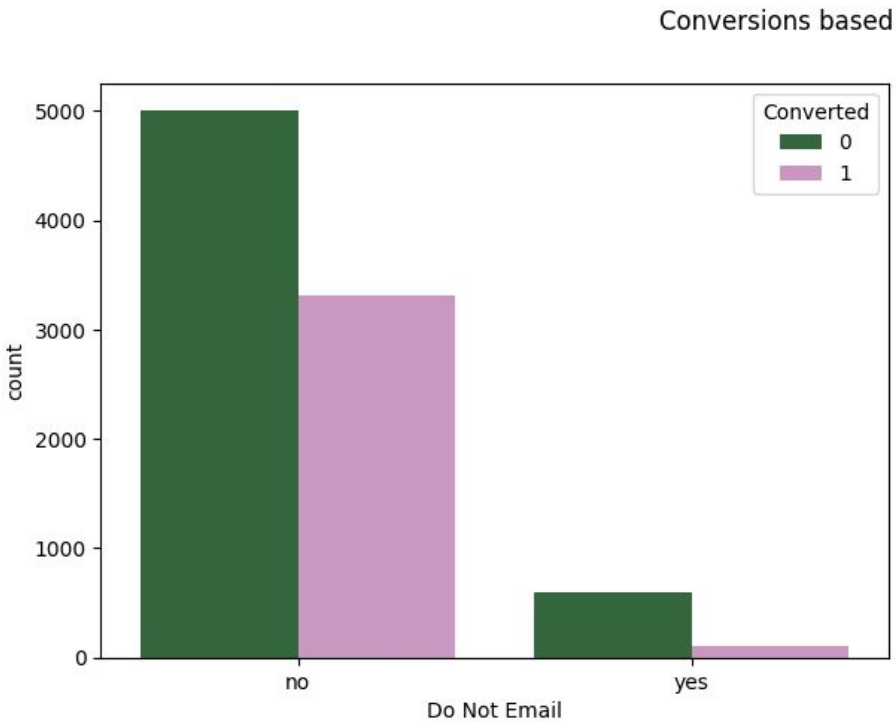
Data Imbalance

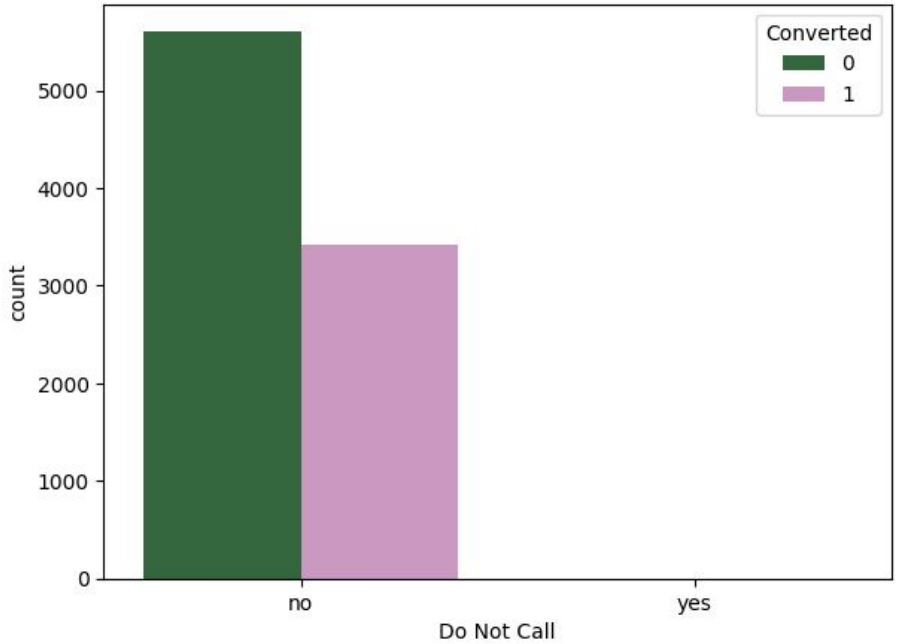




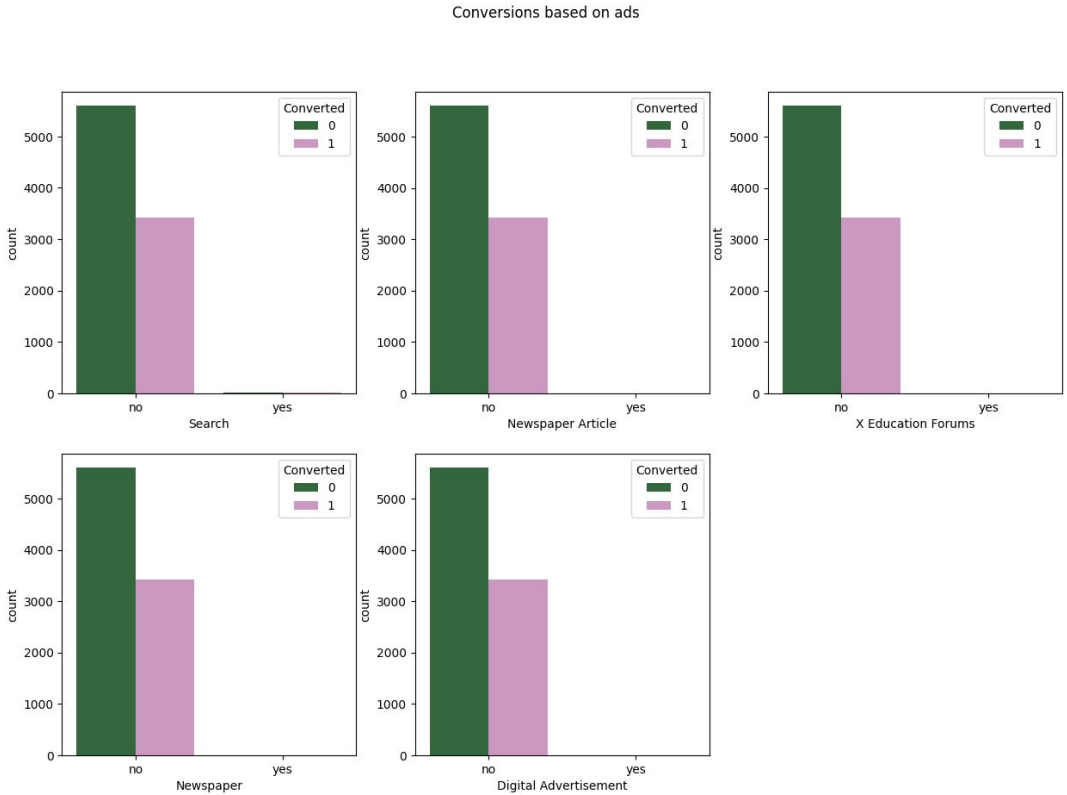
we can say that the data is mildly imbalanced.

Let’s check the Categorical Variables

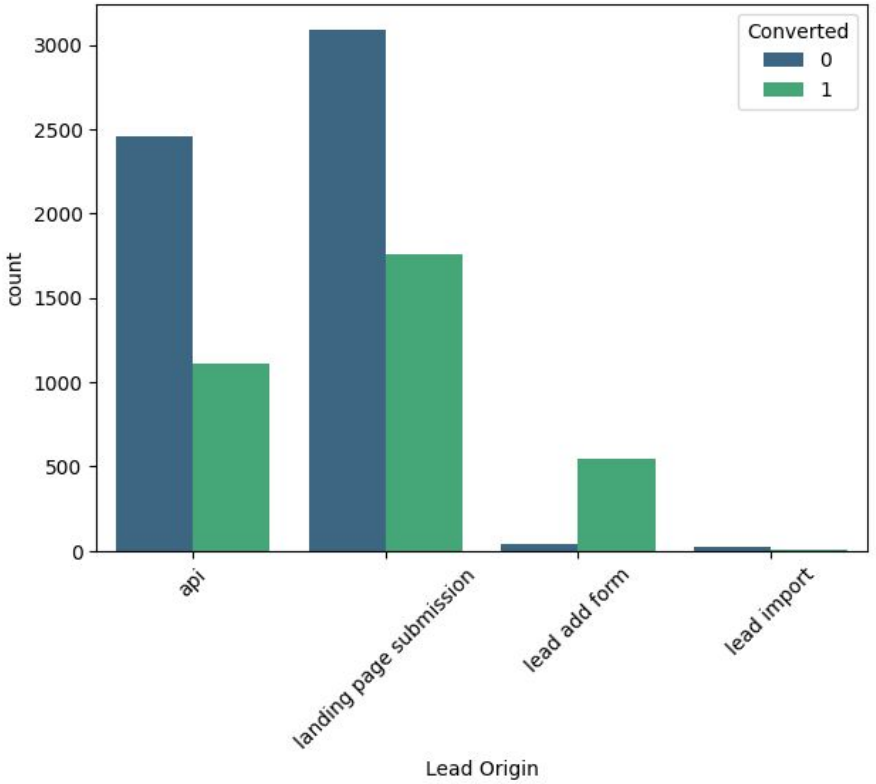


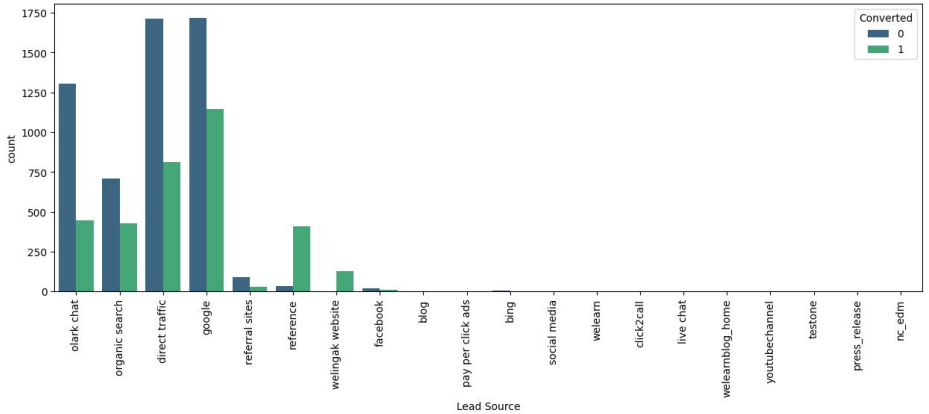


Conversions based on ads

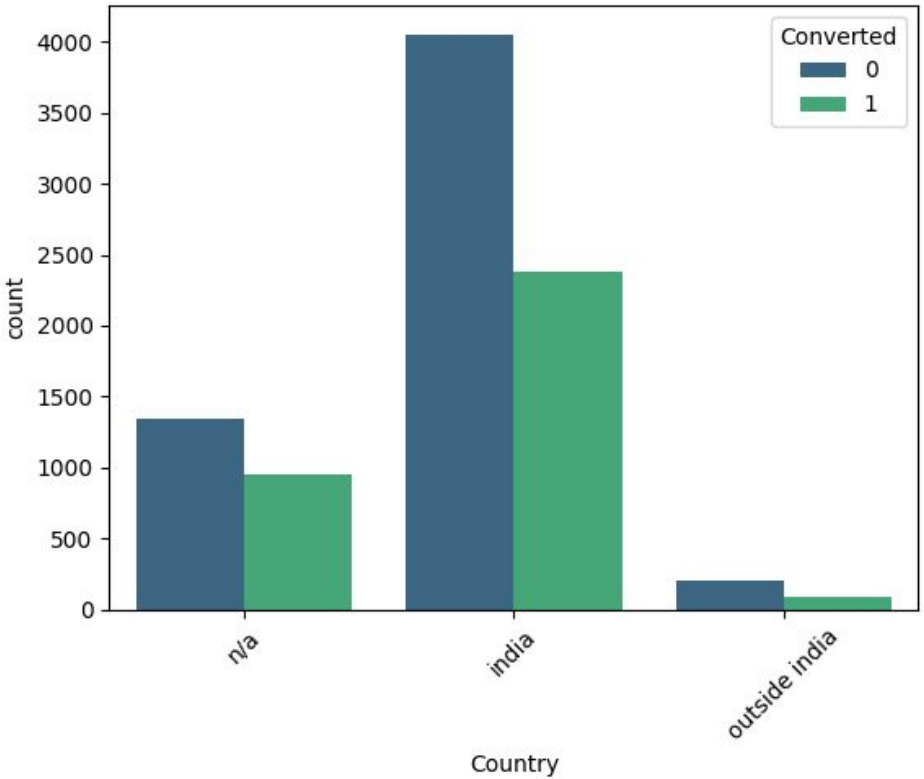


visualizing count of Lead Origin and Lead Source Variable based on Converted value

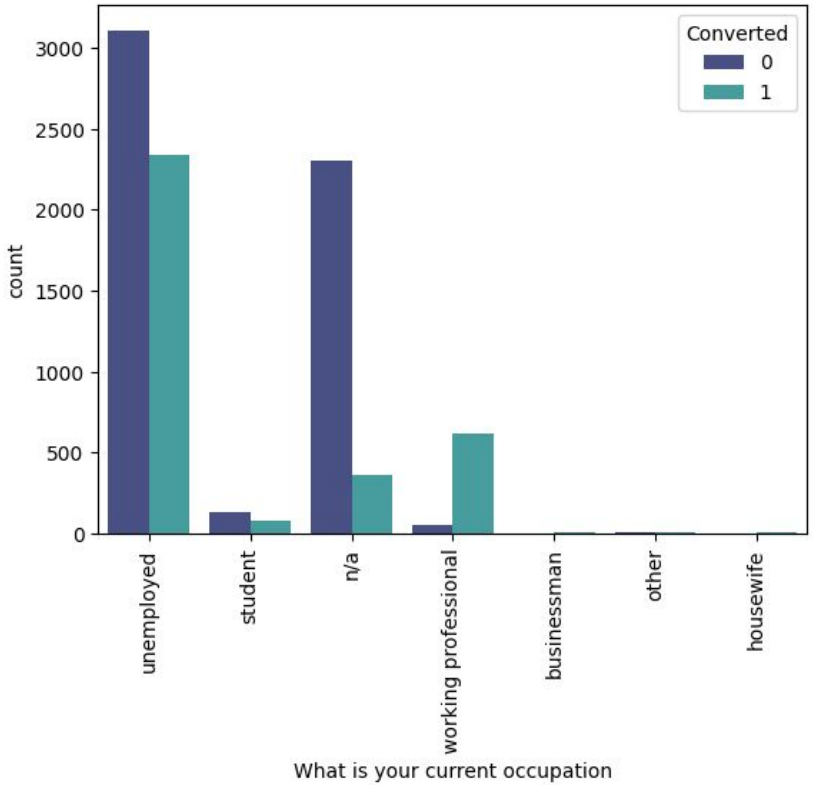




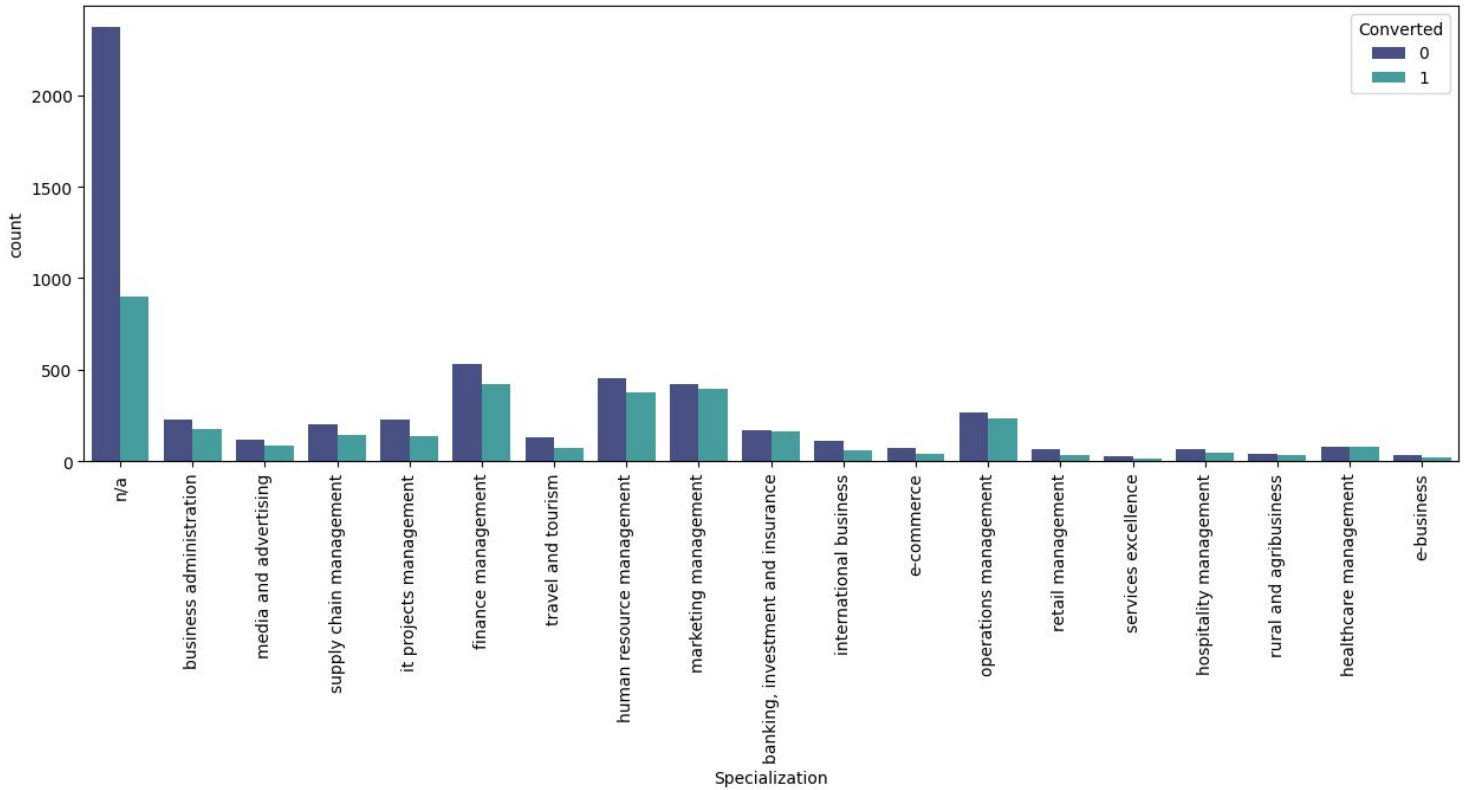
visualizing count of Country Variable based on Converted value



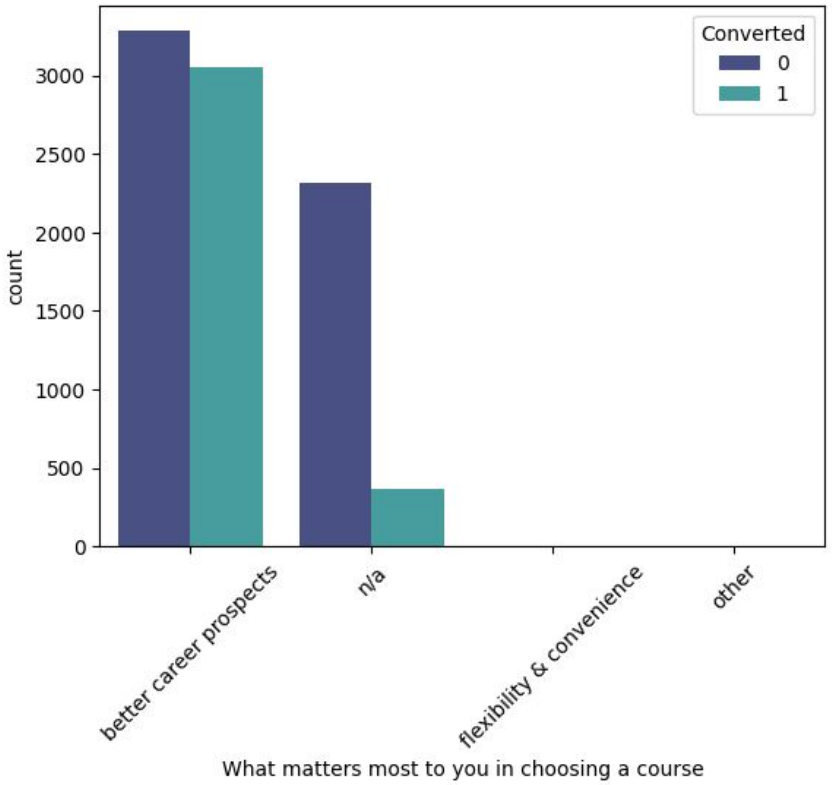
visualizing count of 'What is your current occupation' Variable based on Converted value



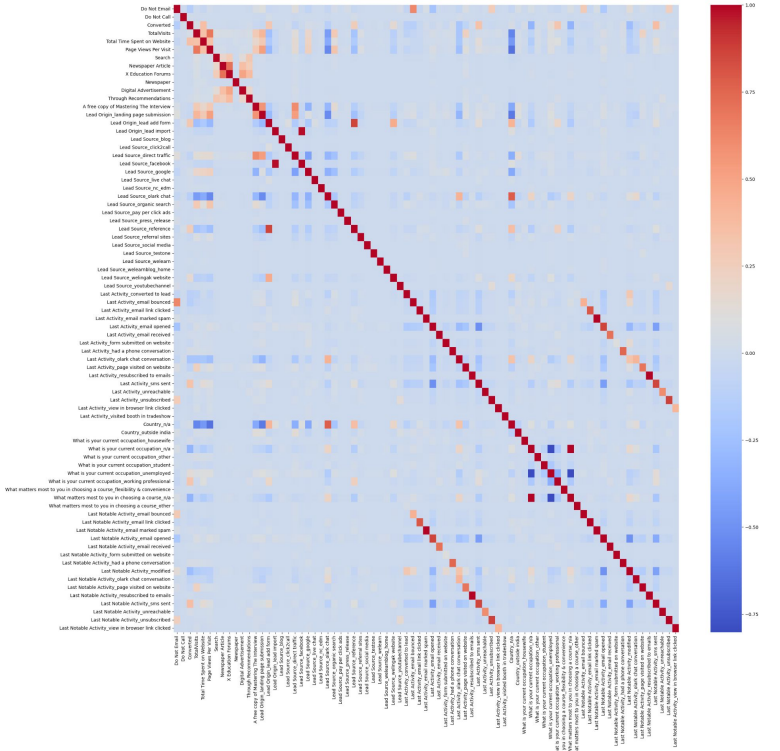
#visualizing count of 'Specialization' Variable based on Converted value



visualizing count of 'What matters most to you in choosing a course' Variable based on Converted value

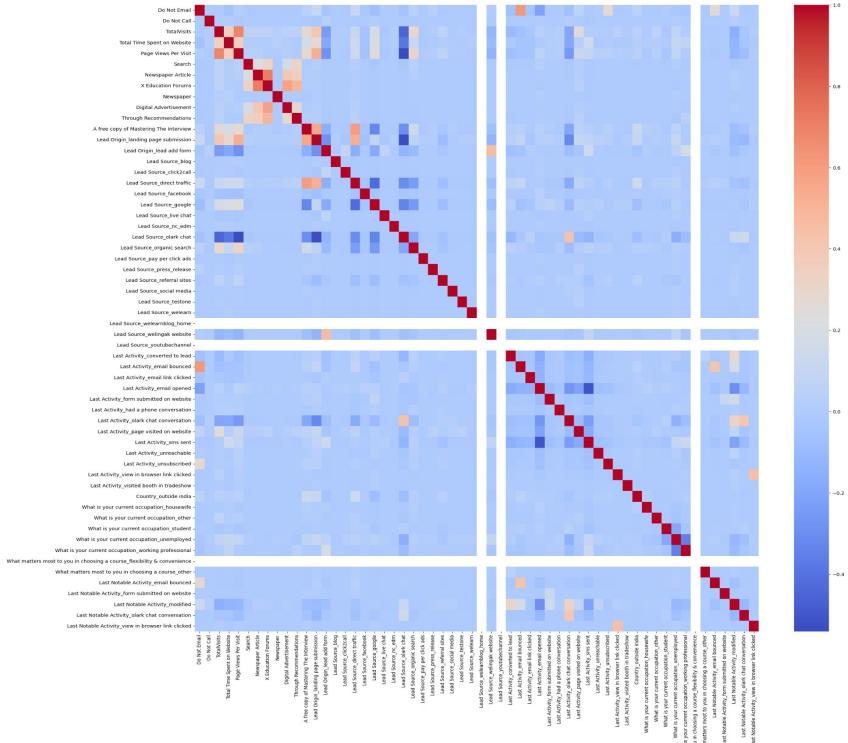


Now Handling Correlation



Due to the large number of variables, correlated variable elimination has to be done programmatically with manual input wherever required.

Checking the correlation again after dropping highly correlated variables



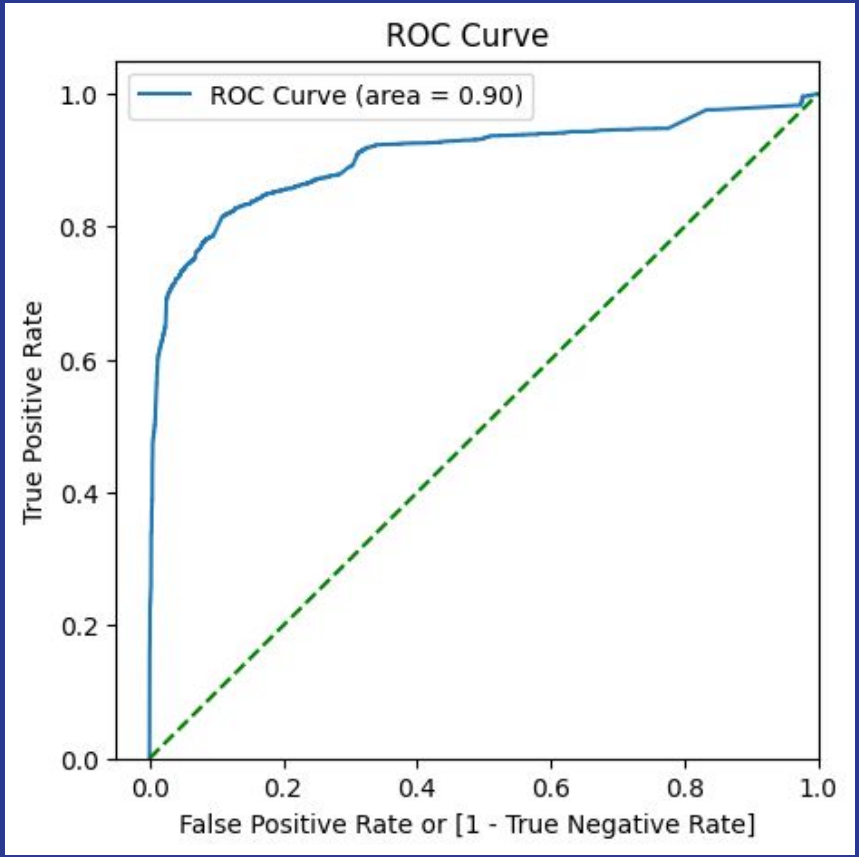
The highly correlated columns have been removed. Now we can proceed further.

After Model Building and Model Evaluation

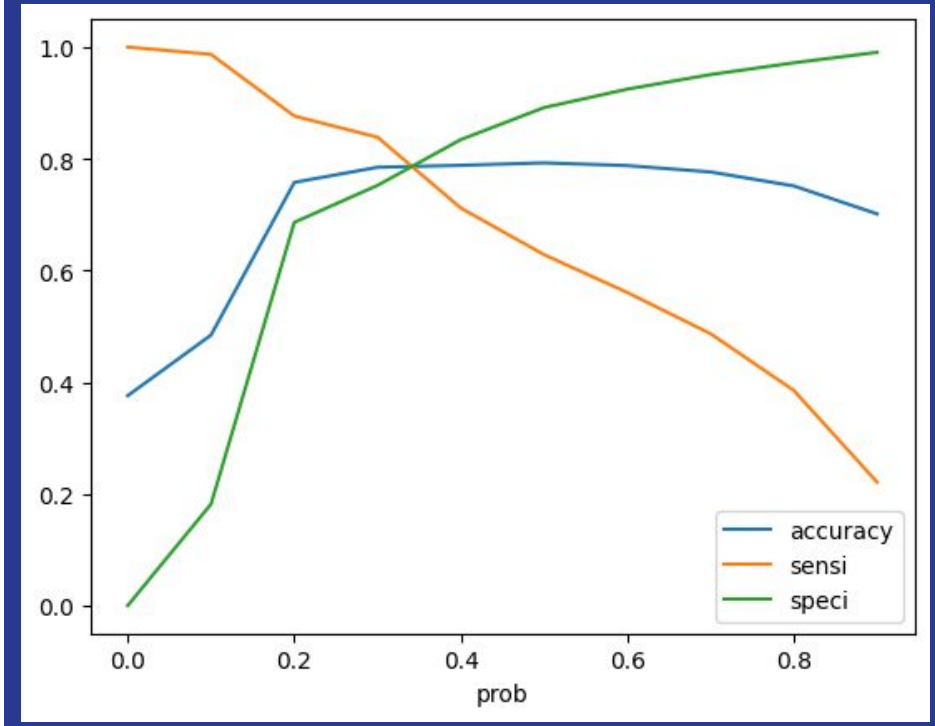
ROC Curve

Finding Optimal Cut off Point

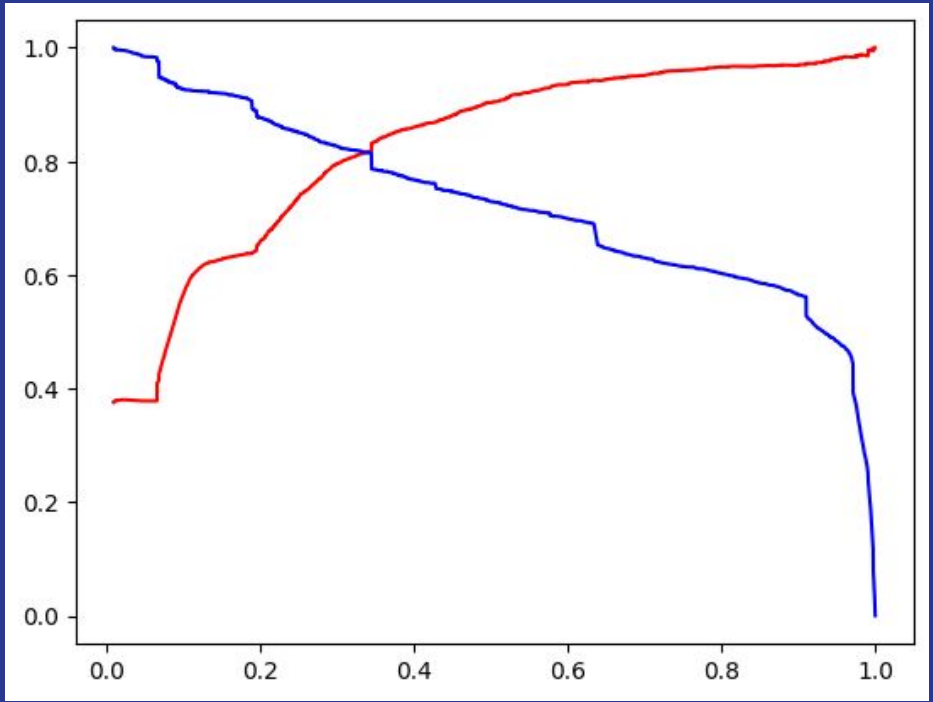
- Optimal cut off probability is that
- probability where we get balanced sensitivity and specificity.



Let's plot accuracy, sensitivity, and specificity for different probability cutoffs



Precision and recall tradeoff



Train set:

Overall accuracy: 0.86

Precision: 0.9

Sensitivity / Recall: 0.73

Specificity: 0.95

Optimal Cutoff Point:

Overall accuracy: 0.85

Precision: 0.8

Sensitivity / Recall: 0.83

Specificity: 0.87

Test set:

Overall accuracy: 0.87

Precision: 0.92

Sensitivity / Recall: 0.73

Specificity: 0.95

Conclusion

To boost course sales, X Education should focus on key factors influencing potential buyers:

Website Engagement:Enhance the online experience to capture buyer interest.

Repeat Visits:Encourage and monitor return visits to maximize conversion chances.

Lead Source Prioritization: Target marketing efforts on high-yield sources like Google, direct traffic, organic search, and the Welingak website.

Last Activity Insight:Utilize data on last interactions (SMS or Olark chat) for timely and relevant follow-ups.

Lead Origin Emphasis (Lead Add Format): Prioritize leads generated through the effective Lead Add format.

Appealing to Working Professionals:Tailor offerings to the needs of working professionals, increasing their likelihood to enroll.

By addressing these factors, X Education can optimize its approach and significantly increase the conversion potential of potential buyers.

X Education's analysis aimed to optimize strategies for attracting industry professionals. A logistic regression study identified key factors: higher platform engagement time, interest in career-focused courses (especially Finance Management), and specializations in HR, Finance, and Marketing Management correlating with increased conversion probability.

To enhance data quality, 'Select' columns were made compulsory, including Customer Occupation and Specialization. Customer engagement through emails and calls, particularly for Finance Management leads, is crucial. Geographical data standardization and exploratory data analysis confirmed no outliers.

Model building via Recursive Feature Elimination yielded six relevant variables with VIF < 5 and p-value < 0.05. Evaluation metrics determined a 0.3 cutoff for around 85% accuracy, sensitivity, specificity, 80% precision, and 82% recall.

Critical variables influencing potential buyers include total time on the website, visits, lead sources, last activity, lead origin, and current occupation as a working professional. Leveraging these insights can significantly enhance X Education's success in attracting and converting potential buyers, guiding targeted marketing and engagement strategies. Precision-recall analysis underscores the importance of a nuanced approach for lead conversion.

Thank You !