Lead Score Assignment

upGrad

Introduction

The Challenge of to convert the conversion rate 30% to 80%.

The assignment involved identifying inefficiencies in the lead conversion process at X Education and proposing strategies to optimize it. X Education is an online education provider that markets its courses across various digital platforms, including websites and search engines like Google. The main challenge identified was the poor conversion rate, with only 30% of acquired leads being converted into customers. This presents an opportunity to enhance the overall lead conversion process by identifying "Hot Leads," or those with the highest likelihood of conversion.

After leads are acquired, the sales team reaches out via phone calls and emails to convert them into customers. However, due to the high volume of leads and the broad targeting strategy, only a small proportion of leads are successfully converted.

How we can address issue

To address these challenges, weare utilizing Exploratory Data Analysis (EDA) techniques. By analyzing lead score application data, we aim to uncover hidden patterns and insights that will help us:

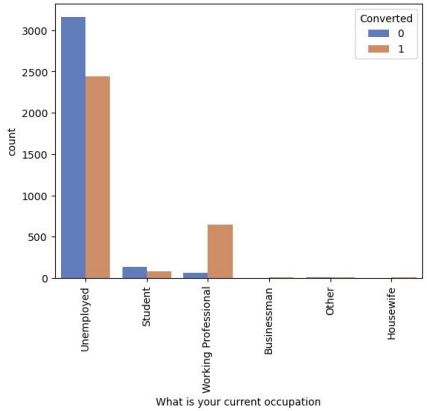
1. **Data understanding, preparation and EDA(Data Analysis and Lead Segmentation)**
2. **Model building and evaluation .**
3. **Final Prioritizing Sales Outreach.**

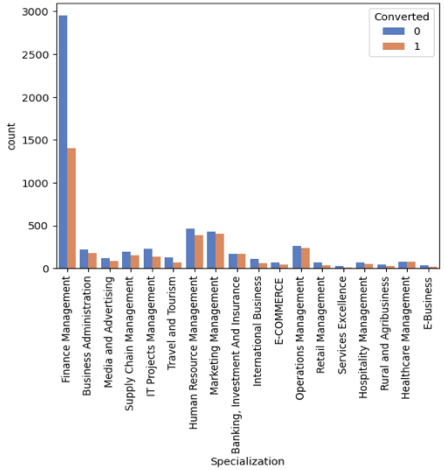
Assumptions

Our approach using **LEAD SCORE CASE STUDY** on lead scoring data relies on several key assumptions:

1. Data Quality: We assume the loan application data is accurate, complete, and representative of the target population (urban loan applicants). Inconsistencies, missing values, or biases in the data can lead to misleading results.
2. Correlation vs. Causation: The analysis will identify correlations between variables and loan defaults. However,correlation doesn't necessarily imply causation. Further investigation might be needed to understand the underlying reasons behind observed patterns.
3. Stationarity of Data: We assume the patterns observed in the historical data will hold true for future loan applications. However, economic conditions, consumer behavior, and lending regulations can change over time,potentially impacting the validity of the model.
4. Predictive Power of New Features: The EDA process might identify new factors beyond traditional credit scores that influence loan repayment. We assume these features will have sufficient predictive power to improve loan approval decisions.
5. Model Generalizability: The model developed through EDA might not be universally applicable to all urban loan applicants. There could be sub-groups within the population with unique characteristics requiring further analysis.
6. Ethical Considerations: We need to be mindful of potential biases in the data or the model itself that could lead to unfair loan approval decisions. Techniques to mitigate bias and ensure fair lending practic

Insights from EDA





1. Current Occupation vs. Conversion:

•Unemployed: Most applicants fall into this category but have a lower conversion rate.
•Working Professionals: Fewer in number but exhibit a higher likelihood of conversion.

2. **Specialization vs. Conversion:**

•Dominant Specializations: Finance Management and Business Administration have the highest number of applicants.
•Conversion Trends: Some specializations consistently convert at higher rates, indicating targeted interest areas.

Data Cleaning and Processing

Checked the null data.

In [11]:
```
# Check the number of null values again
leads.isnull().sum().sort_values(ascending=False)
# same as 9
```

Out[11]:

Lead Quality	4767
Asymmetrique Activity Index	4218
Asymmetrique Profile Score	4218
Asymmetrique Activity Score	4218
Asymmetrique Profile Index	4218
Tags	3353
Lead Profile	2709
What matters most to you in choosing a course	2709
What is your current occupation	2690
Country	2461
How did you hear about X Education	2207
Specialization	1438
City	1420
Page Views Per Visit	137
TotalVisits	137
Last Activity	103
Lead Source	36
Receive More Updates About Our Courses	0
I agree to pay the amount through cheque	0
Get updates on DM Content	0
Update me on Supply Chain Content	0
A free copy of Mastering The Interview	0
Prospect ID	0

Model Building

Test-Train Split

The next step is to split the dataset into training an testing sets.

	Prospect ID	Lead Number	Lead Origin	Lead Source	Do Not Email	Do Not Call	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Last Activity	...	Lead Quality	Update me on Supply Chain Content	Get updates on DM Content	Asymmetrique Activity Index	Asymm Profil
0	7927b2df-8bba-4d29-b9a2-b6e0beafe620	660737	API	Olark Chat	No	No	0.0	0	0.0	Page Visited on Website	...	Low in Relevance	No	No	02.Medium	02.I
1	2a272436-5132-4136-86fa-dcc88c88f482	660728	API	Organic Search	No	No	5.0	674	2.5	Email Opened	...	NaN	No	No	02.Medium	02.I
2	8cc8c611-a219-4f35-ad23-fdfd2656bd8a	660727	Landing Page Submission	Direct Traffic	No	No	2.0	1532	2.0	Email Opened	...	Might be	No	No	02.Medium	
3	0cc2df48-7cf4-4e39-9de9-19797f9b38cc	660719	Landing Page Submission	Direct Traffic	No	No	1.0	305	1.0	Unreachable	...	Not Sure	No	No	02.Medium	
4	3256f628-e534-4826-	660681	Landing Page	Google	No	No	2.0	1428	1.0	Converted	...	Might be	No	No	02.Medium	

Insights from Logistic Regression Model

Key Determinants of Lead Conversion:

1.Engagement Metrics:

1. Total Visits: High-frequency visitors are significantly more likely to convert.
2. Total Time Spent on Website: Longer time spent indicates higher interest and conversion probability.

1.Lead Sources:

1. Lead Add Form: Captures high-quality leads, indicating serious inquiries.
2. Olark Chat: Real-time assistance enhances conversion likelihood.
3. Welingak Website: An effective lead generation source.

1.Communication Preferences:

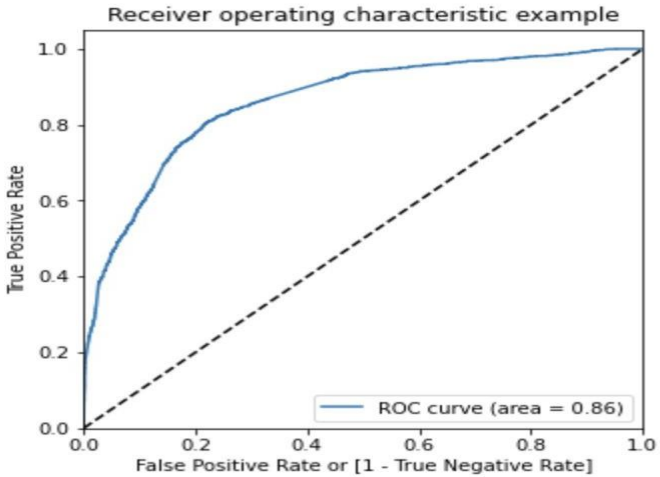
1. Do Not Email (Yes): Opt-out leads are less likely to convert, highlighting the importance of email marketing.
2. Phone Conversations & SMS: Personal and direct communication positively impacts conversions.

1.Lead Characteristics:

1. Occupation (Student/Unemployed): Lower conversion likelihood due to financial constraints.
2. Last Notable Activity (Unreachable): Unreachable leads may still convert through follow-ups.

Model Evaluation

Finding the Optimal Cutoff

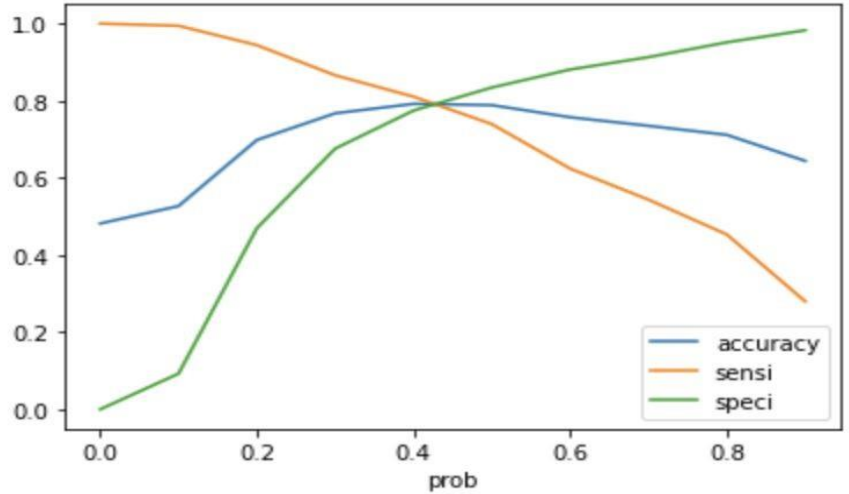


The area under the curve of the ROC is 0.86 which is quite good. So we seem to have a good model. Let's also check the sensitivity and specificity tradeoff to find the optimal cutoff point.

Model Evaluation

Plot cutoff value

```
plt.show()
```



As you can see that around 0.42, you get the optimal values of the three metrics. So let's choose 0.42 as our cutoff now.

Suggestions based on Findings

Boost Website Engagement: Add engaging content and re-engage visitors effectively.

Optimize Lead Forms: Improve form design and offer incentives to increase submissions.

Leverage Chat & Calls: Enhance chat availability and train staff for better interactions.

Run Targeted Campaigns: Personalize emails and address specific lead preferences.

Focus on Key Segments: Tailor strategies for low-converting groups or prioritize others.

Refine Lead Follow-Ups: Analyze and improve processes for "unreachable" leads.

Conclusion

In conclusion, by identifying "Hot Leads," X Education can optimize its lead conversion process, improve its efficiency, and potentially increase its overall revenue. This project underscored the value of combining data analysis with strategic decision-making in sales and marketing.

Future Scope

Enhanced Predictive Modeling Using Advanced AI/ML Techniques:

- **Incorporating more sophisticated machine learning models** such as deep learning, decision trees, or ensemble methods like Random Forest or XGBoost could improve prediction accuracy. These models can handle more complex data relationships and provide better classification of leads.
- **Real-time prediction models** can be developed to score leads dynamically as they interact with the website or other platforms. This would allow the sales team to instantly identify high-potential leads and act promptly.
- **Natural Language Processing (NLP)** can be integrated to analyze communication patterns between leads and the sales team, extracting insights from emails or chats to assess a lead's likelihood of conversion.