

Lead Scoring Case Study

Improving Lead Conversion for X Education

Predictive model for Identifying Hot Leads

Submitted by -
Dhruva Handa,
Dhwani Hingorani,
Mohit Dhawan





Problem Statement

- **Current Scenario**

- X Education markets online courses to industry professionals.
- Visitors land on the website through various channels (websites, search engines, referrals).
- Visitors can browse courses, fill out forms, or watch videos.
- Leads are generated when visitors provide their email or phone number.
- Sales team contacts leads via calls and emails.
- **Current conversion rate: ~30%.**

- **Challenges**

- High volume of leads with low conversion efficiency.
- Inefficient allocation of sales team efforts.



Goals to Achieve

- **Objective**
 - Improve the lead conversion rate for X Education by identifying and prioritizing "Hot Leads" that are most likely to convert into paying customers, aiming for a target lead conversion rate of 80%.
- **Solution Approach**
 - Develop a predictive model to assign a lead score to each lead.
 - Focus sales efforts on leads with higher scores to increase conversion rates.
- **Expected Outcome**
 - Improved lead conversion rate.
 - More efficient sales team operations.



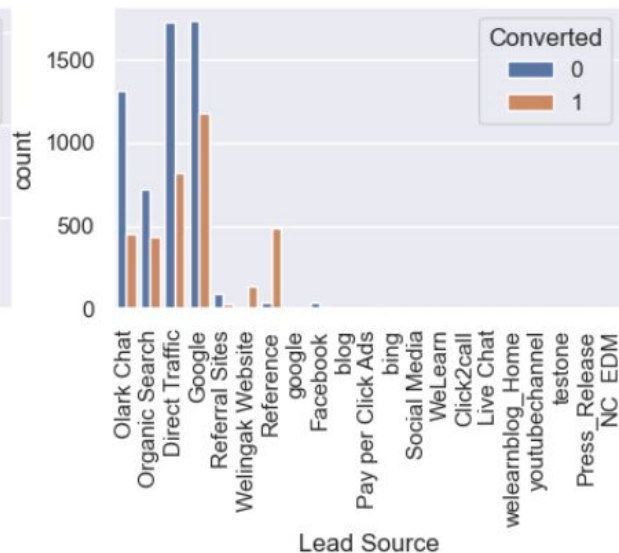
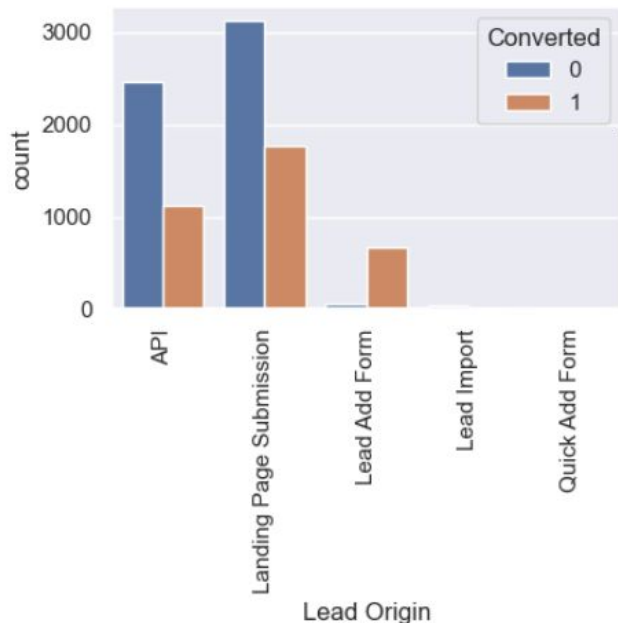
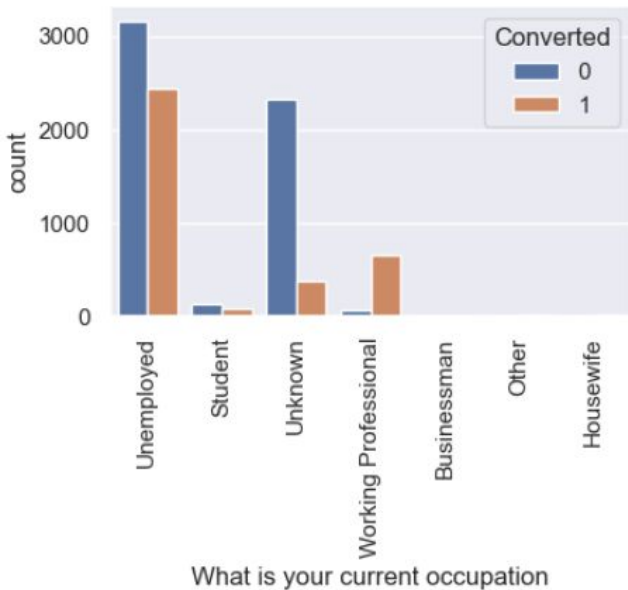
Strategy - Analysis Approach

- **Data Cleaning and Preparation**
- **Feature Engineering and Scaling**
- **EDA - Exploratory Data Analysis**
- **Model Building**
 - Model Type: **Logistic Regression**
 - Used Recursive Feature Elimination to shortlist top 20 features and then did manual fine tuning using VIF and p-values to further optimize the model.
- **Model Validation**
 - Used validation metrics like Accuracy, Sensitivity, Specificity and predictive values.
 - Determining optimal cut-off value using these metrics.
 - Testing on Test Data using the optimal cut-off obtained.

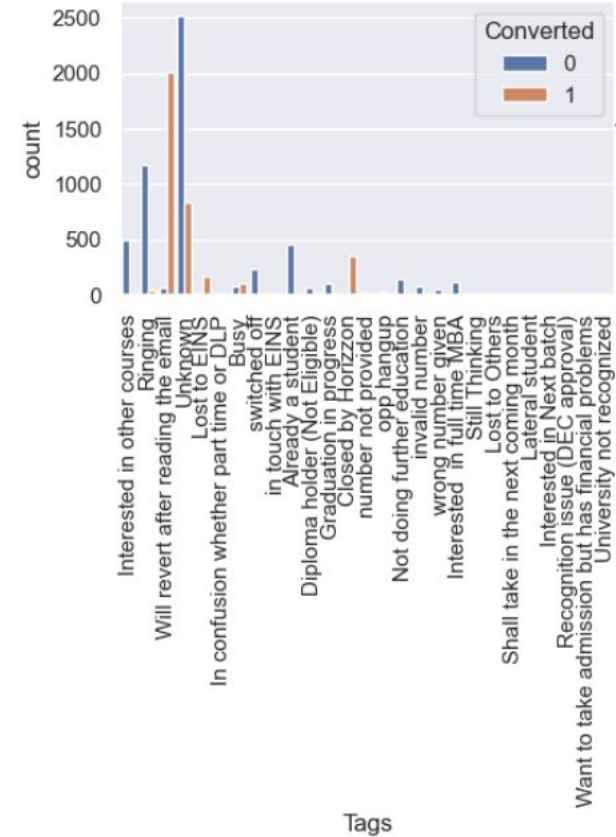
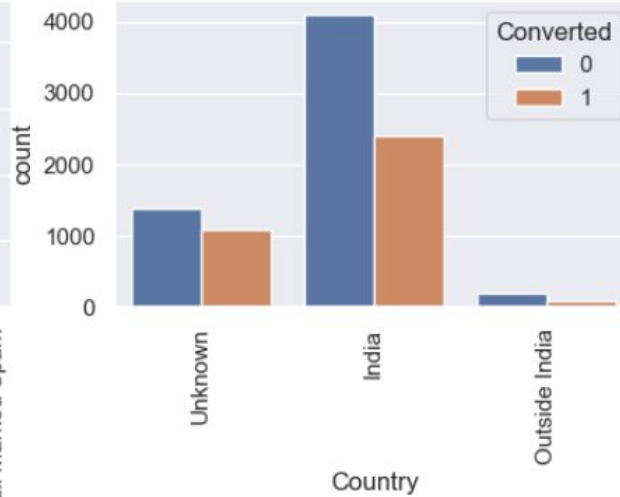
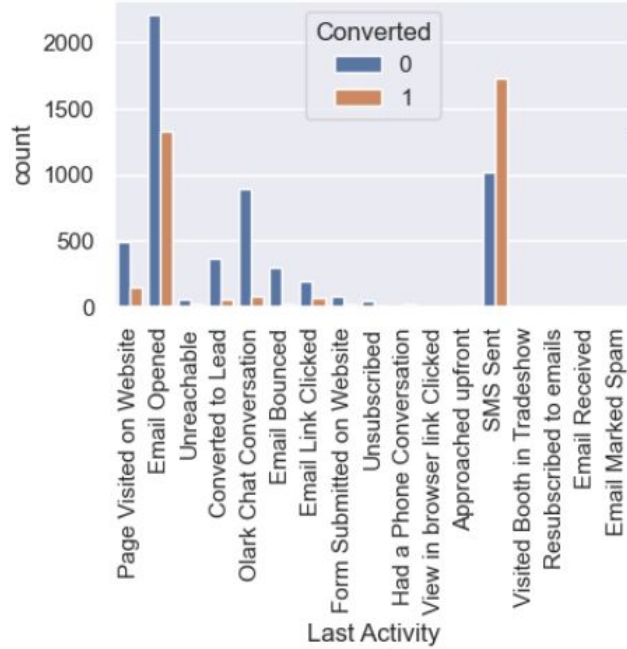


Exploratory Data Analysis

Here are some graphs for reference



Exploratory Data Analysis -II





Variables impacting the conversion rate

Lead
Source_Wellin
gak_website

Last
Activity_SMS
Sent

What is your
current
occupation_Un
employed

What is your
current
occupation_Wo
rking
Professional

Tags_Closed by
Horizzon

Last Notable
Activity_Email
Link Clicked

Tags_Intereste
d in full time
MBA

Tags_Intereste
d in other
courses

Tags_lost to
EINS

Tags_Not doing
further
education

Tags_Ringing

Last Notable
Activity_Modified

Tags_Will
revert after
reading the
email

Tags_invalid
number

Tags_opp
hangup

Tags_switched
off

Lead
Quality_Worst



Model Evaluation: Accuracy and Sensitivity on Train and Test Dataset

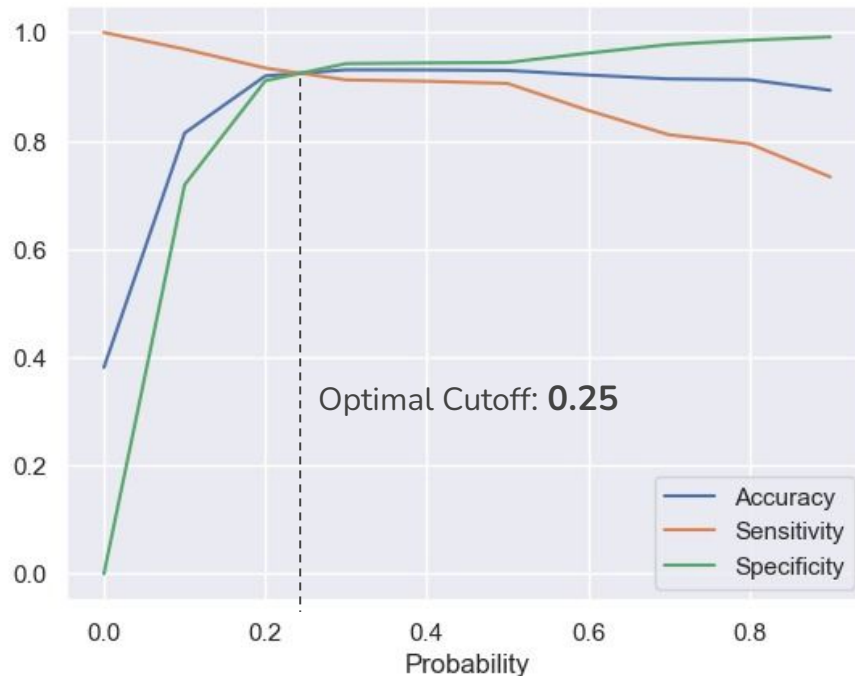
With this cut-off, the final Model has achieved the following metrics:

- Accuracy: 91.9%
- Sensitivity: 93.4%

Using the same cutoff on the Test data, we still managed to achieve:

- Accuracy: 92.2%
- Sensitivity: 94.8%

Hence, we have a good model.





Conclusion and Recommendations

Key Takeaways:

- Before Model Implementation, **Conversion rate** was 30% and after Model Implementation it comes out to be **>90%** on both Train and Test sets.
- By focusing on leads with higher scores, the **sales team** can prioritize efforts on the most promising leads, thus **increasing overall efficiency**.
- The logistic regression model demonstrated **high Accuracy, Sensitivity, and Specificity**, indicating reliable performance across both the training and test sets.

Recommendations:

- **Continuous Model refinement** with new data to ensure it adapts to changing trends and behaviors.
- Conduct **training sessions to help the sales team** understand and effectively use lead scores in their outreach efforts.
- Implement a feedback loop to continuously **monitor conversion rates** and gather insights for further optimization.
- Explore additional features and data sources to enhance the model's predictive power (e.g., social media engagement, advanced demographic data).



THANK YOU :)

Appendix





Model Evaluation - Precision and Recall on Train dataset

Based on the precision-recall curve, we had obtained an optimal cutoff of 0.41. However, we wanted to proceed with the Accuracy, Sensitivity and Specificity curves, so we have illustrated that approach.

