upGrad

Lead Score Case Study

Shreshtha Doharey

Problem Statement

X Education, an online course provider, attracts industry professionals through various marketing channels like websites and search engines. When interested visitors fill out a form, they are classified as leads. Despite acquiring many leads, only 30% get converted, leading to an inefficient process. To improve conversion rates, the company aims to identify 'Hot Leads'—those with the highest potential to convert—allowing the sales team to focus their efforts more effectively and improve overall lead conversion. The goal is to optimize the middle stage of the funnel by nurturing high-potential leads.

Objective

The objective of this project is to build a predictive model for X Education to assign a lead score to each potential customer, identifying those most likely to convert into paying customers. The model should prioritize leads based on their conversion likelihood, enabling the sales team to focus on high-potential leads ('Hot Leads'). The goal is to increase the overall lead conversion rate to approximately 80%, as outlined by the CEO. This will enhance efficiency in the lead conversion process by targeting the most promising prospects.

To Solve Problem Approach Followed

1. Importing the data and inspecting the data frame
2. Data cleaning
3. Data visualization
4. Data Preprocessing (creating Dummy variables and scaling)
5. Data Modeling (using RFE , P-value and VIF)
6. Data Evaluation
7. Optimal cut Off selection
8. Prediction on the test set

Data Manipulation

✓Total number of row =9240 and columns =37

✓Columns dropped with null valus greater than 40% .

✓Dropped single variable columns like 'Magazine','Receive More Updates About Our Courses','I agree to pay the amount through cheque‘ .

✓ Also dropped two variable column with maximum weight age to only variable are of no use in prediction like ‘Do not Email’ and ‘Do not call’.

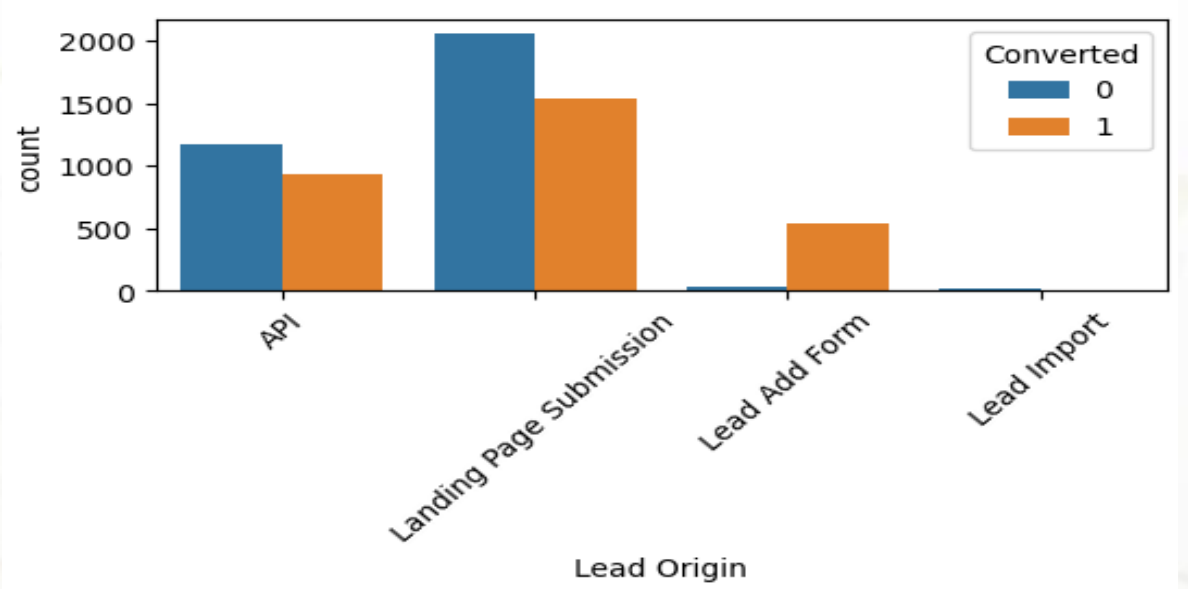
✓ Categorical columns in which some of the variables with nunique count less than 15 are dropped as it will increase the number features which are not even that impactful

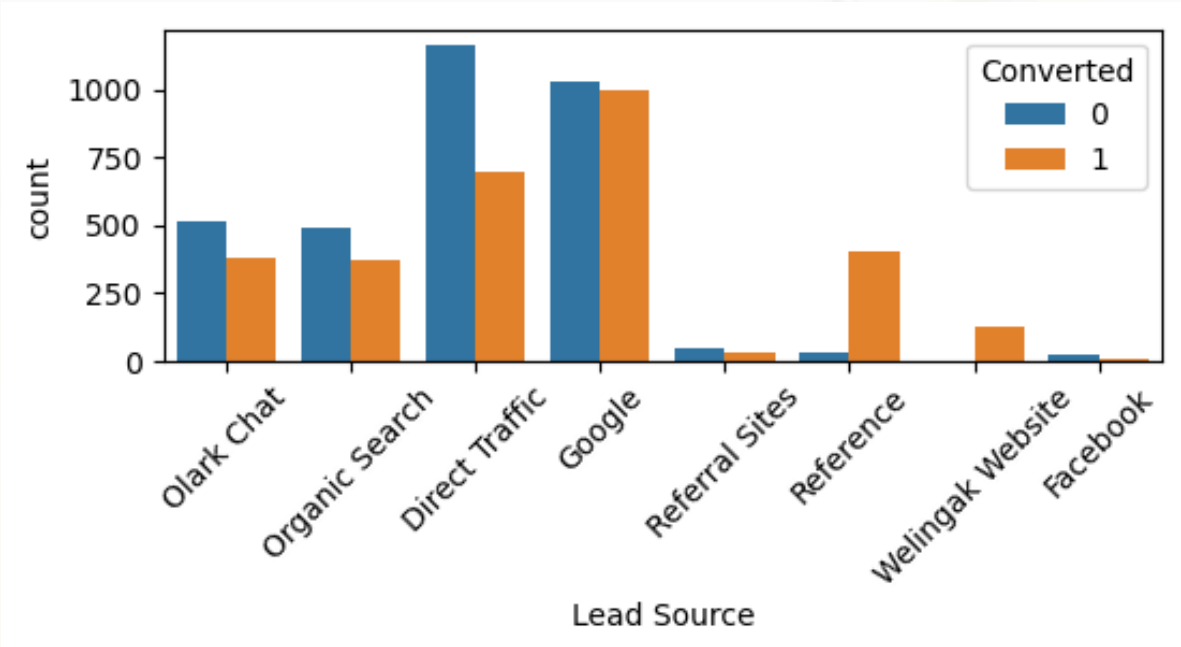
✓Treated outliers using capping method to avoid the data loss.

Data visualization

upGrad

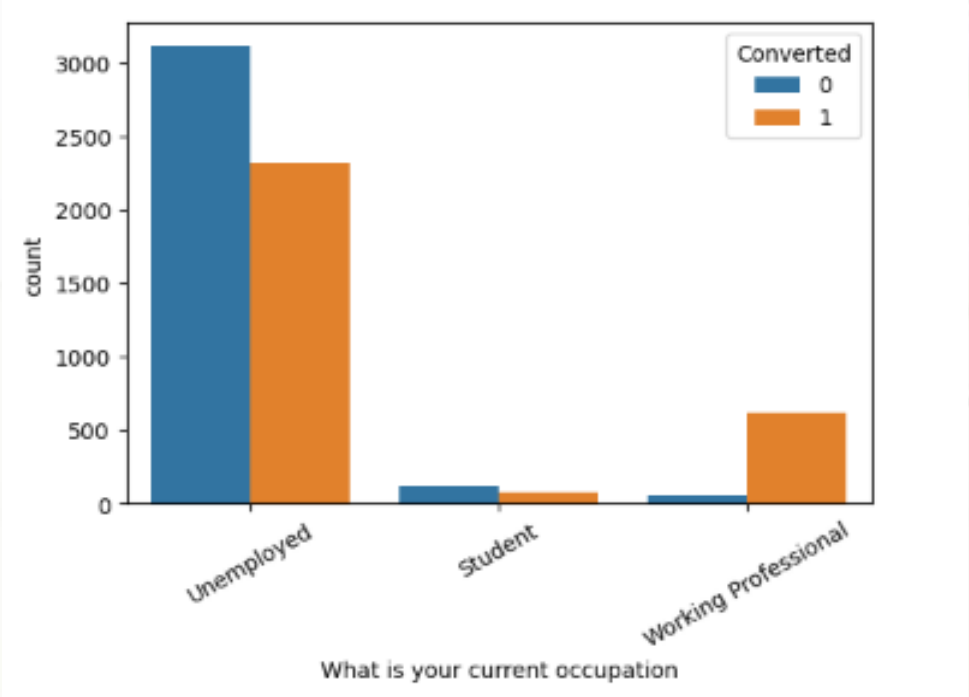
✓In Lead Origin Lead Add Form has shown maximum conversion.





✓Lead Sources like Google , Reference , welingagk Website are the hot leads

Data visualization

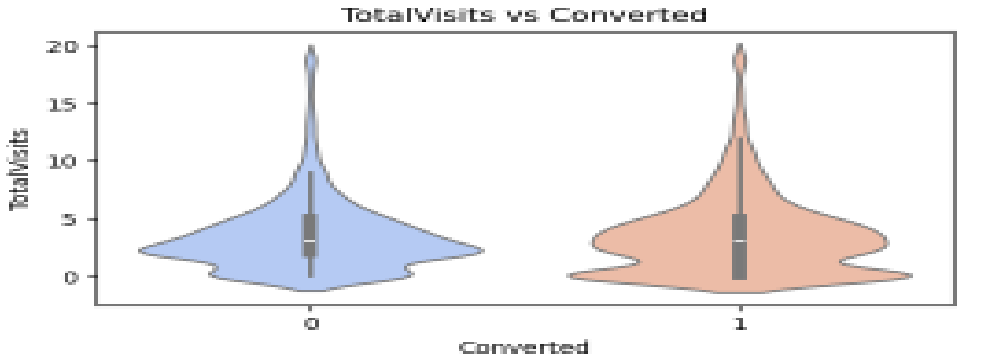


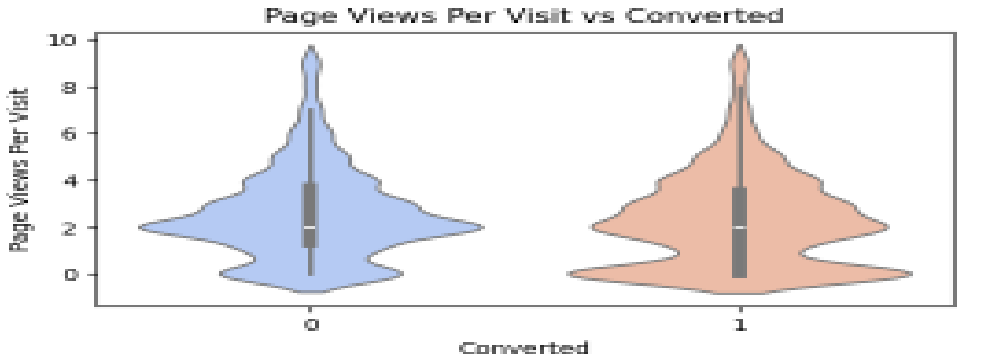
✓Working Professional are more likely to convert

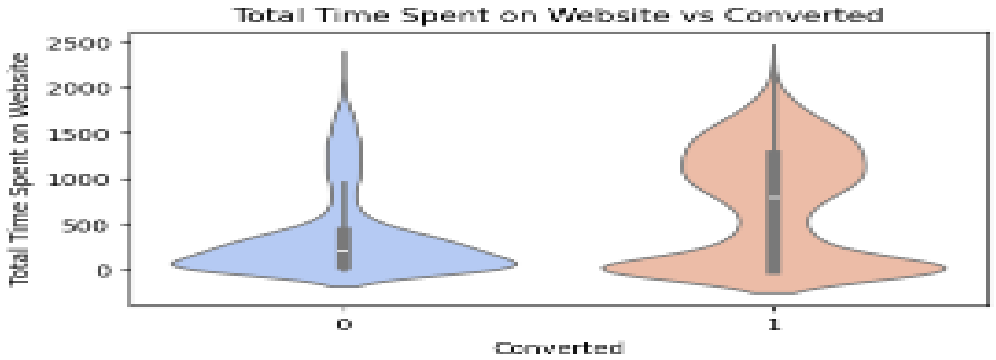
✓Unemployed are not that interested may be because of the financial reasons but still a part to focus .

✓Students are not someone good to invest time and efforts

Numeric Variables







✓As the number of Total Time spent on website increases probability of conversion also increases

✓Total visits and page views per visit increase also impacting the lead decisions.

Data Conversion

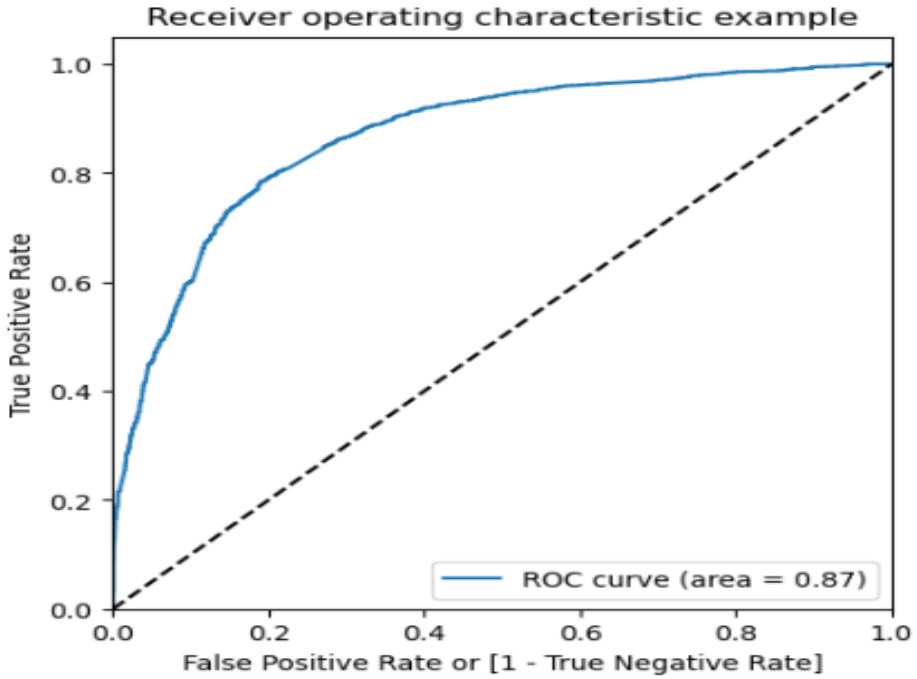
- ✓ Dummy Variable are created from object type variables
- ✓ Numeric columns are normalized using minmax scaler.
- ✓ Total row for analysis 6300
- ✓ Total columns for analysis 51

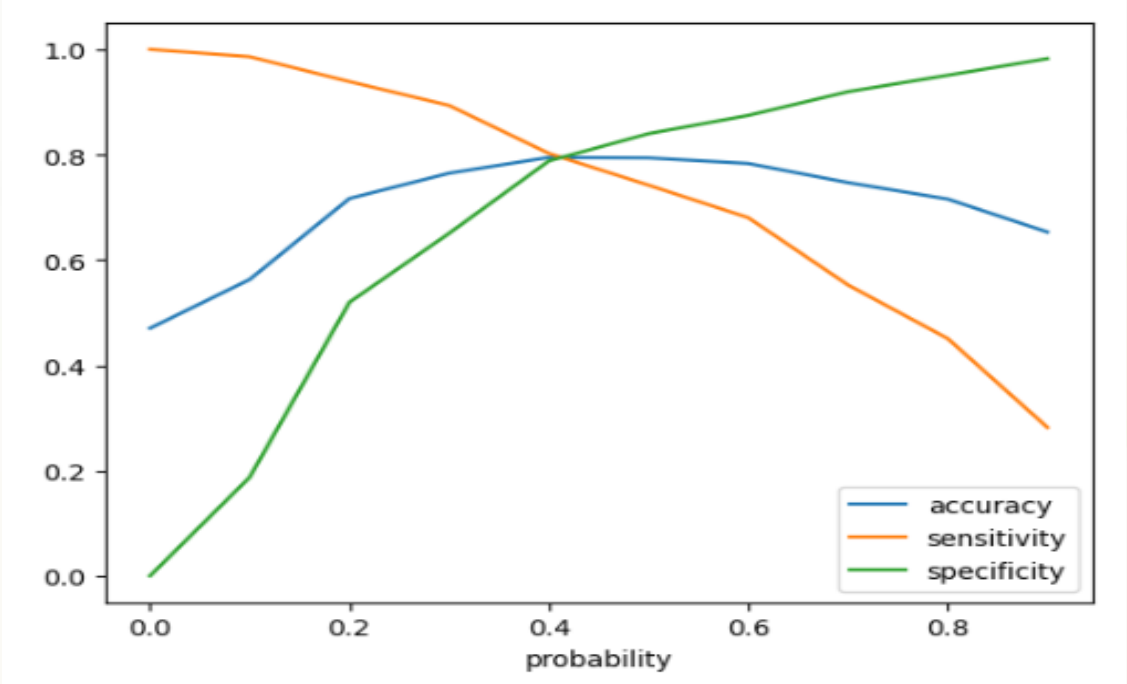
Data Modelling

- ✓ Splitting the data into Training and Test set in the ration 70:30
- ✓ Used RFE for feature selection
- ✓ 15 Features selected with the RFE
- ✓ Manually Tuning the 15 features by checking P-value less than 0.05 and VIF not greater than 5
- ✓ 12 features obtained after building 4 models
- ✓ Obtain optimal cut off 0.42 and Roc curve 0.87
- ✓ Predicted the test set
- ✓ Accuracy obtain 78%

ROC Curve

- ✓ Area Under the ROC curve is 0.87 which is close to represents the strong discriminationg power of the model .





- ✓ Optimal cut-off obtained through the range of probablitiy cutoff is 0.42

Optimal Cutoff selection

Key Insights:

✓Feature Impact: The analysis revealed that leads from the Lead Add Form had a significantly higher likelihood of conversion.

✓Furthermore, Google, Reference, and the Welingak Website emerged as top-performing sources, demonstrating high conversion rates.

✓Activity-Based Engagement: Leads showing activities such as SMS Sent and Email Opened were more engaged and had a higher chance of conversion. This highlighted the importance of these activities as key indicators of lead interest.

✓Specialization Significance: Certain specializations such as Business Administration, Human Resource Management, Marketing Management, Banking, Investment and Insurance, Finance, and Operation Management were identified as high-potential areas. Leads in these fields were found to be more likely to convert.

✓Target Audience: Working Professionals were noted as a particularly interested demographic. Tailoring marketing efforts to this group and emphasizing the flexibility and benefits of the offerings could lead to higher enrollment rates.

Conclusion:

✓Enhance Visibility of the Lead Add Form: Improve placement and accessibility to capture more high-converting leads.

✓Optimize Marketing on Key Sources: Allocate resources to enhance visibility on Google, Reference, and the Welingak Website.

✓Focus on High-Interest Activities: Implement targeted follow-up strategies for leads with SMS Sent and Email Opened.

✓Target Hot Specializations: Customize marketing and outreach efforts for the identified high-potential specializations.

✓Address Working Professionals: Develop specialized programs and highlight benefits that align with the career needs of working professionals.