

LEAD SCORING CASE STUDY

Problem Statement

- ❖ X Education, an education company, sells online courses to industry professionals.
- ❖ Many professionals land on their website and browse for courses on a daily basis. The people landing on their website may either browse the courses, watch some videos or fill up a form for a course.
- ❖ On filling the form with their email address or phone number, they are classified as a lead.
- ❖ However only some of these leads get converted into paying customers. The typical conversion rate at X Education is 30%, which is poor as compared to the leads they get.
- ❖ To increase the conversion rate, the company wishes to identify the most potential leads called “Hot Leads”.
- ❖ This identification can help the company’s sales team to focus on communicating more with these potential leads rather than calling everyone.

Business Objective

- ❖ Identify the most promising lead, i.e. the ones who will convert into paying customers.
- ❖ Build a model to identify these potential leads.
- ❖ Try to achieve a lead conversion rate of around 80% as expected by the CEO of X Education by identifying the hot leads correctly.

Approach

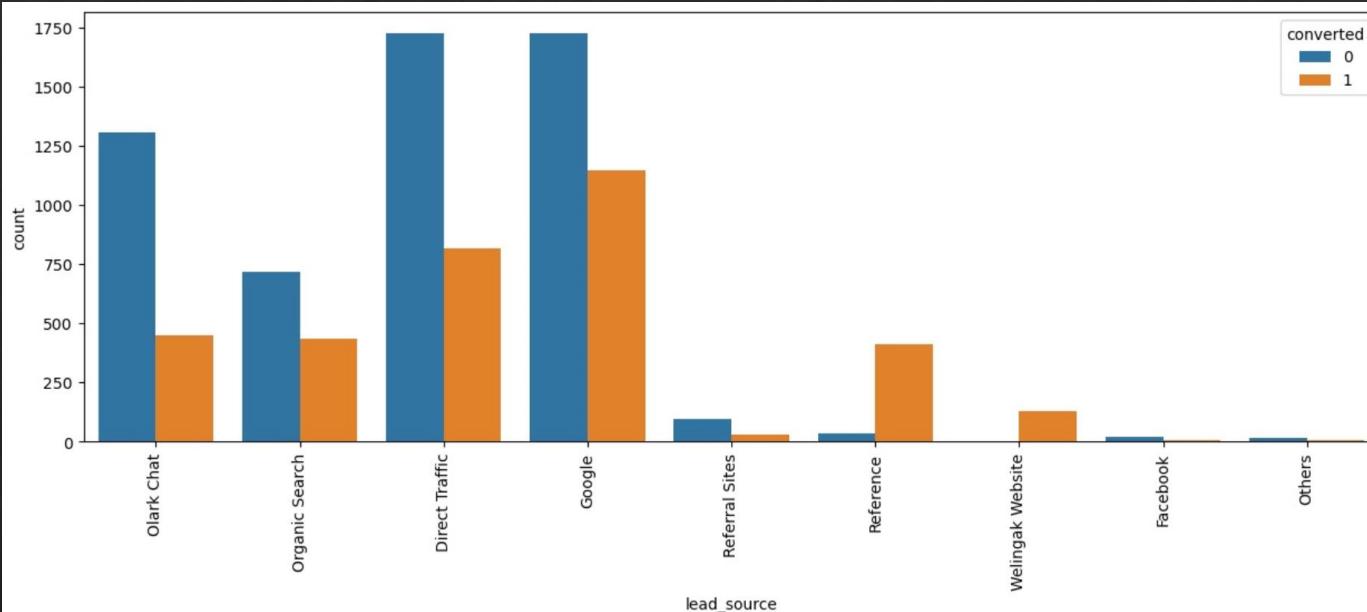
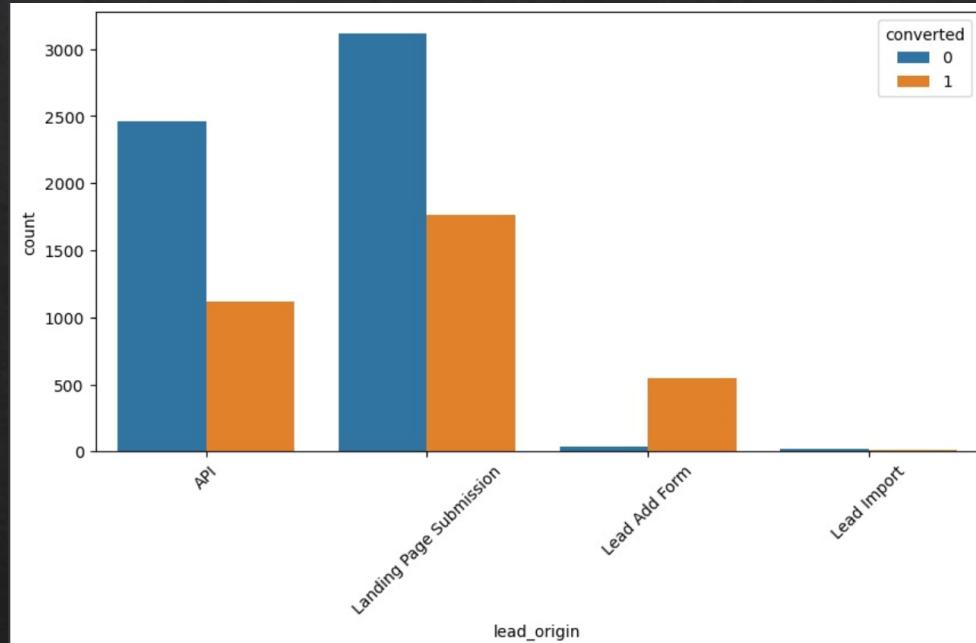
- ❖ Read and understand the data
- ❖ Data preprocessing
 - i. Handle null/missing values
 - ii. Handle outliers
- ❖ Exploratory Data Analysis (EDA) – Univariate and Bivariate analysis
- ❖ Data Preparation
 - i. Dummy variable creation
 - ii. Test-Train split
 - iii. Feature Scaling
- ❖ Model Building
- ❖ Model Evaluation
- ❖ Make predictions on test data

Data Preprocessing

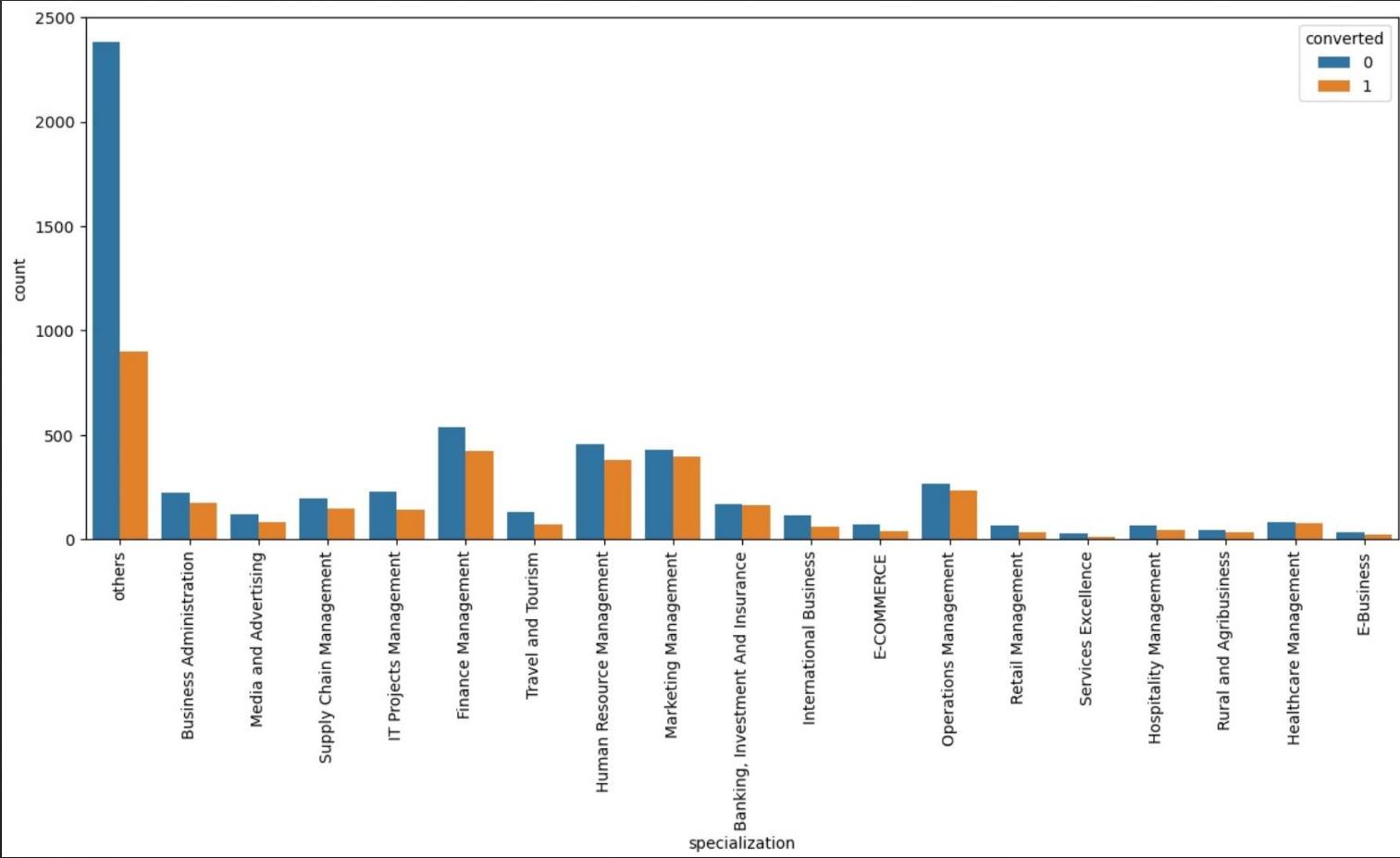
- ❖ Renamed the long feature names to short ones.
- ❖ Checked for duplicates and also checked for columns with ‘Select’ label.
- ❖ Four features had the ‘Select’ category which was replaced with null value.
- ❖ Deleted some features with more than 40% missing values and for some features we decided to impute with either the mode or create a new category.
- ❖ Some features had less than 1% missing values so dropped those missing rows.
- ❖ Dropped some columns which were not important/necessary for the analysis.
- ❖ Also handled the outliers.

Exploratory Data Analysis (EDA)

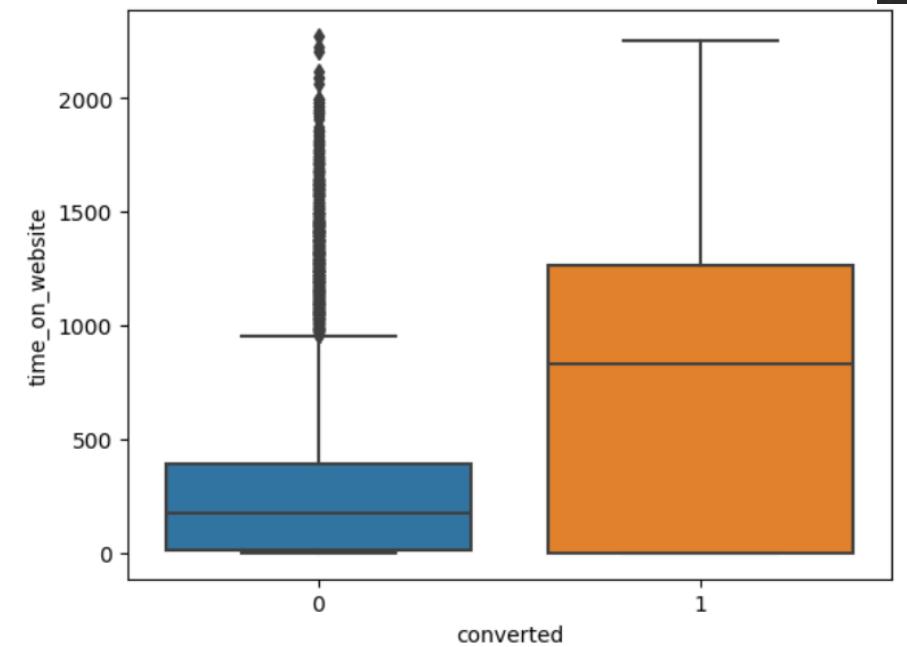
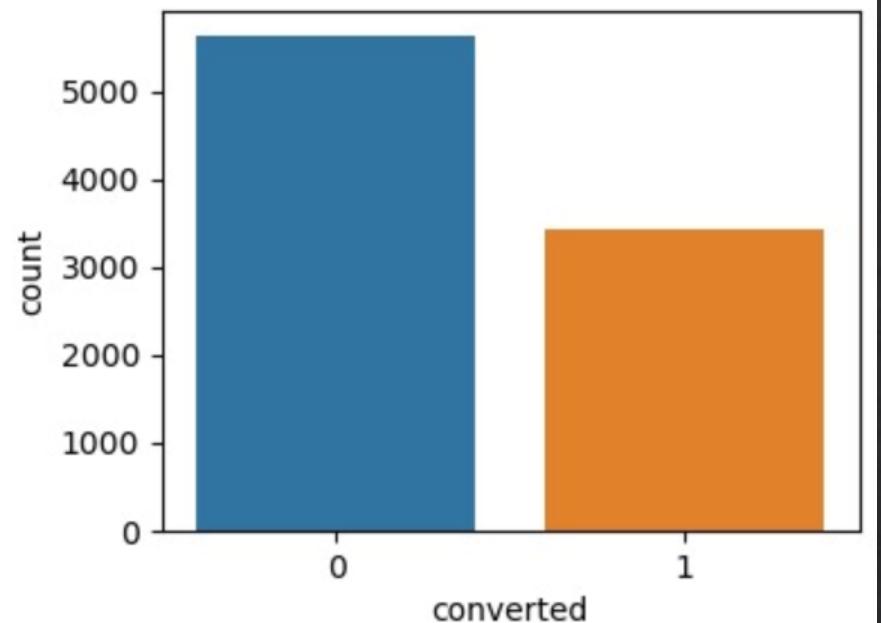
- ❖ We performed univariate and bivariate analysis on the features.



- Google and Direct Traffic generates maximum leads
- Conversion rates of leads through reference and Welingak website is high.



- There is a poor lead conversion rate. Only 37.85% leads have converted.
- Leads who are spending more time on the website are more likely to be converted.



Data Preparation

- ❖ Converted some binary variables (Yes/No) to 1/0.
- ❖ Created dummy variables for categorical features.
- ❖ Split the data into test and train sets as below:

Train set - 70%

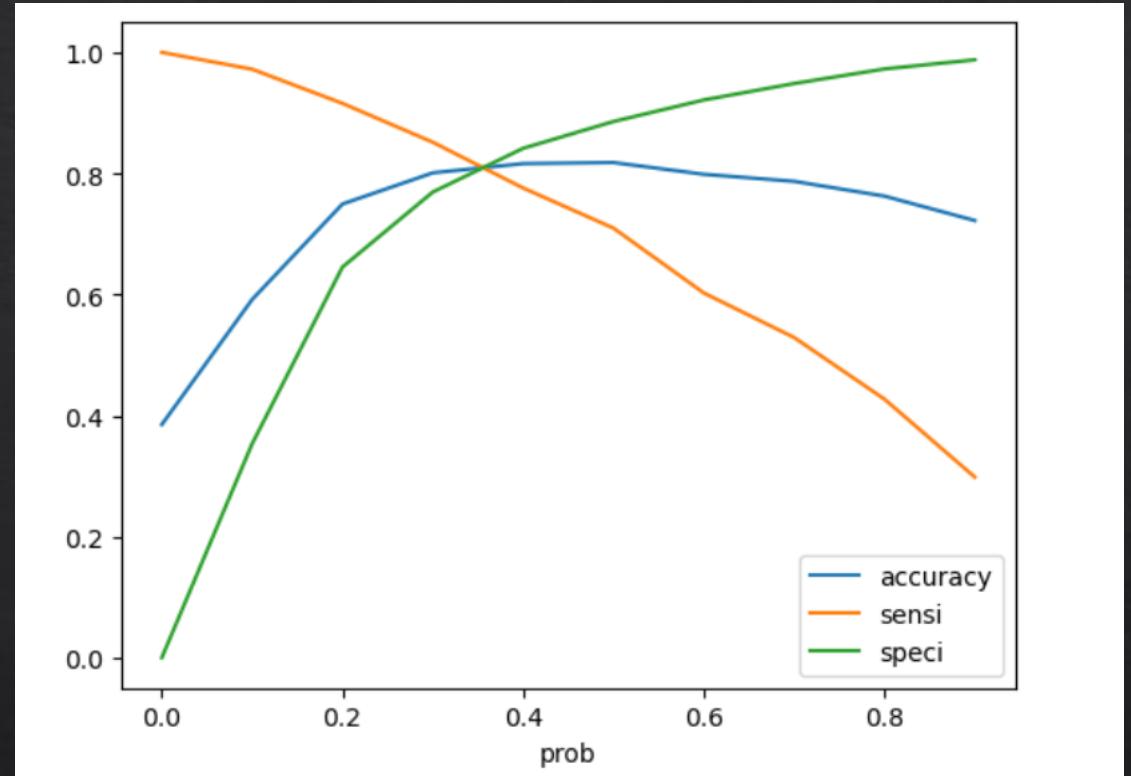
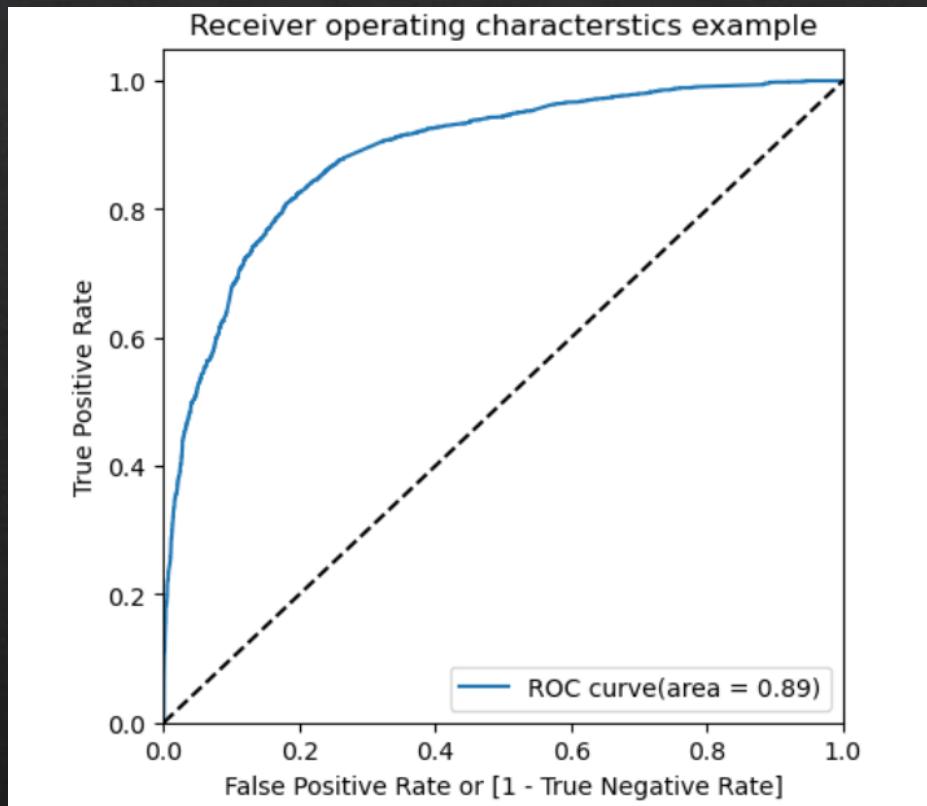
Test set - 30%

- ❖ Scaled features using StandardScaler().

Model Building

- ❖ Selected 20 features using RFE.
- ❖ Reduced features by manually dropping them based on p-value and VIFs.
- ❖ Repeated the process till we got all features with p-value and VIFs within the accepted range (p-value < 0.5, VIF < 5).
- ❖ Model -7 was the final model with 14 features. This was used to make predictions.

Model Evaluation



- Created Confusion matrix.
- Model has a higher area (0.89) under the ROC curve proving it is a good one.
- Plotted accuracy, sensitivity and specificity for various probabilities to find the optimal cutoff probability of 0.34.

Make predictions on test data

- ❖ Scaled the features of test data using StandardScaler().
- ❖ Made predictions on the final model.
- ❖ Assigned lead score.

Observations

- ❖ Below are the results of the accuracy, sensitivity and specificity of the train and test data.
- ❖ Train Data
 - i. Accuracy – 81.0%
 - ii. Sensitivity – 81.9%
 - iii. Specificity – 80.5%
- ❖ Test Data
 - i. Accuracy – 80.4%
 - ii. Sensitivity – 80.4%
 - iii. Specificity – 80.4%

Conclusion

- ❖ Leads coming from the lead sources ‘Welingak Websites’, ‘Reference’ and ‘Olark Chat’ are more likely to get converted.
- ❖ Working professionals are more likely to have a high conversion rate.
- ❖ Leads who spend more time on the websites are likely to get converted.
- ❖ The sales team of X Education should focus on communicating more with these leads.