

Lead Scoring Case Study

By:-

Balaji Shaik

Shiva Kant Singh

Aishwarya Kumar Sharma

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor.
- For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on
- communicating with the potential leads rather than making calls to everyone.

Business Objective

- X Education wants to know most promising leads.
- For that they want to build a Model which identifies the 'Hot Leads'.
- Deployment of the Model for the future use.

Solution Methodology

- For EDA process we are following steps and approach:
 - ❖ **Data Cleaning :**
 - Feature Selection:
 - in this section we find out important column by using some methodology
 - by dropping the those column which have more than 35% null value
 - by dropping the those value which have less corelation with target column
 - **Feature Engineering**
 - Missing Imputation
 - categorical variable: we will impute them
 - with most repeated value. Mode will give us the most repeated value.
 - numerical variable: we will impute
 - with Mean if the data is distributed normally.
 - with Median if there are outlier present.
 - with mode if replacing with most repeated value makes sense.
 - Value Modification in this section we will change data type ,correct format of value etc.
 - Outlier Detection & Treatment
 - for identification most of use boxplot and Scatter plot
 - use quintile for better look on outlier
 - we have treatment outlier by using following technique
 - Imputation of outlier
 - Delete outlier if not impact the analysis
 - Binning(Bucketing)
 - Cap the outlier

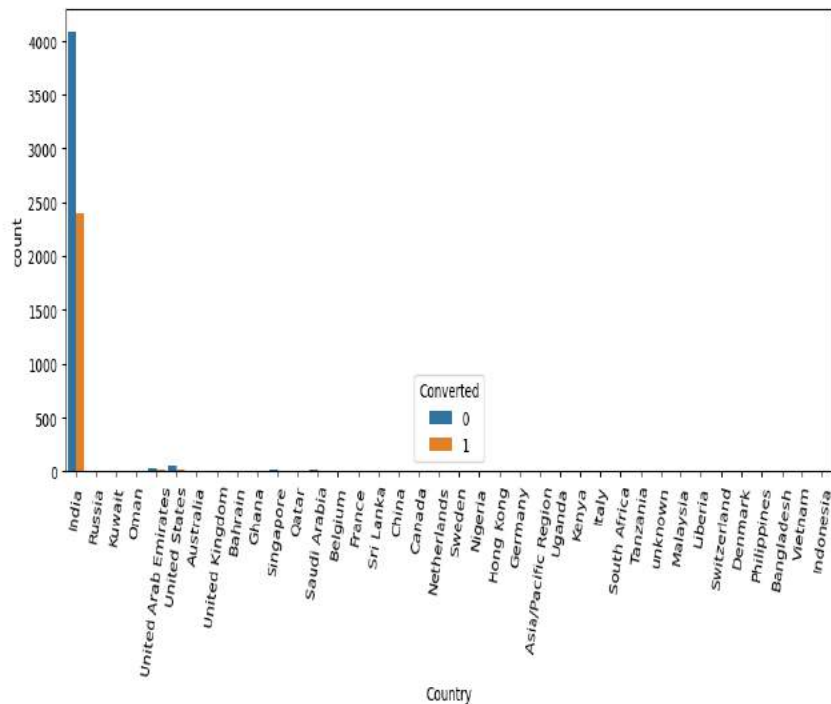
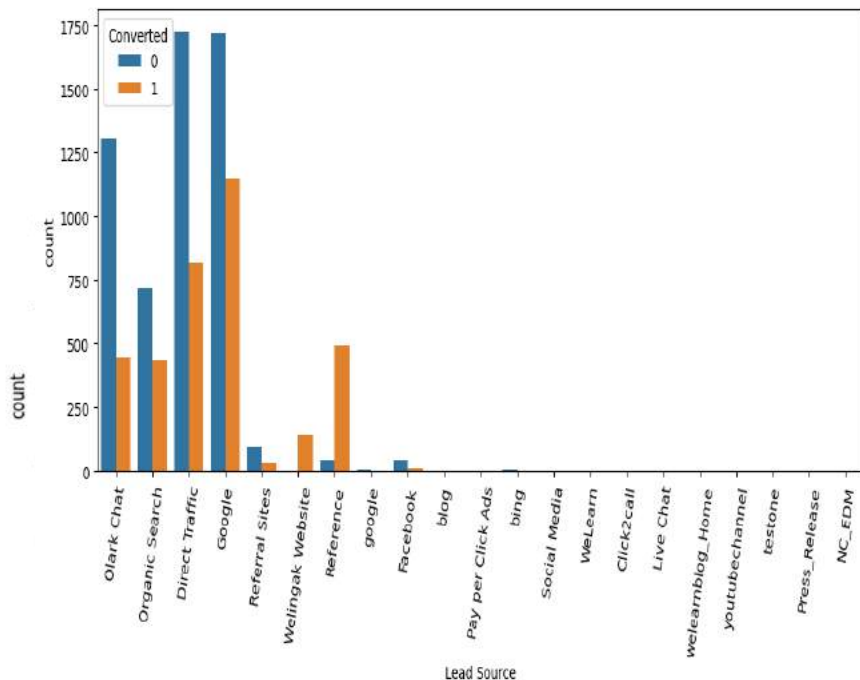
Solution Methodology (Continued)

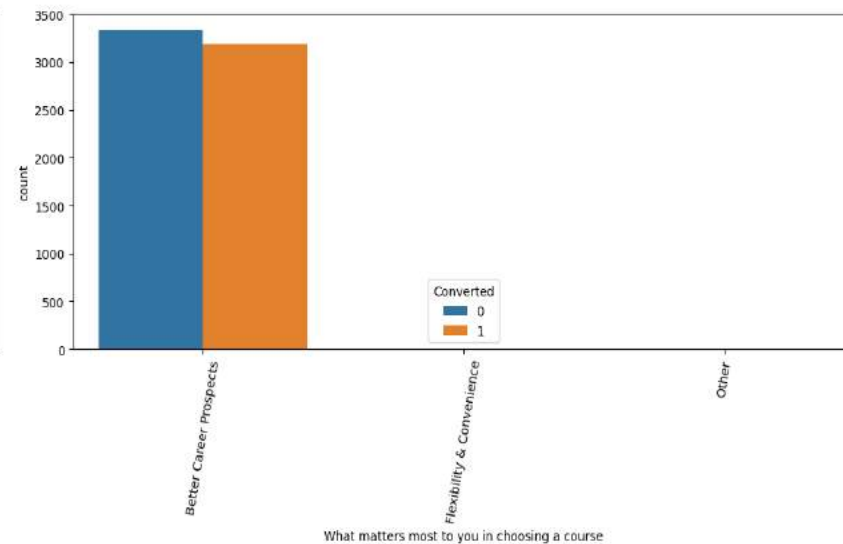
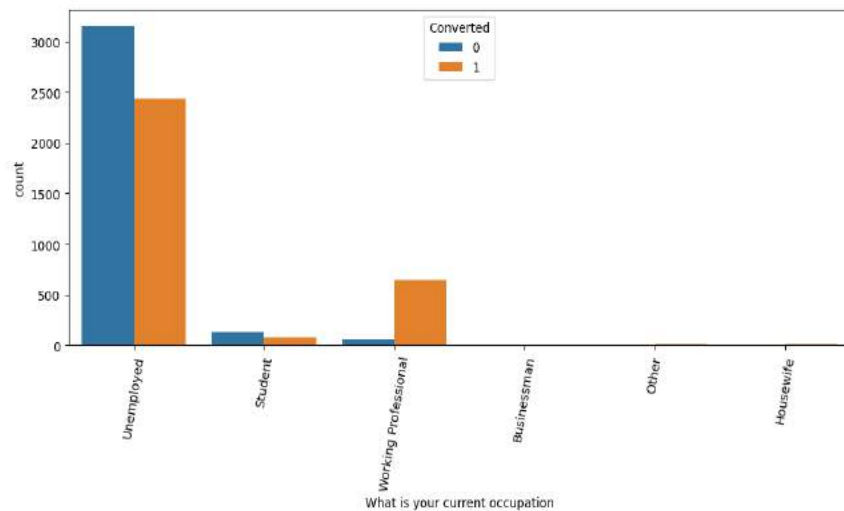
- Feature Scaling and Dummy Variables
- Classification Techniques: Logistic Regression used for the model making and prediction
- Validation of the Model
- Model presentation
- Conclusions and Recommendations

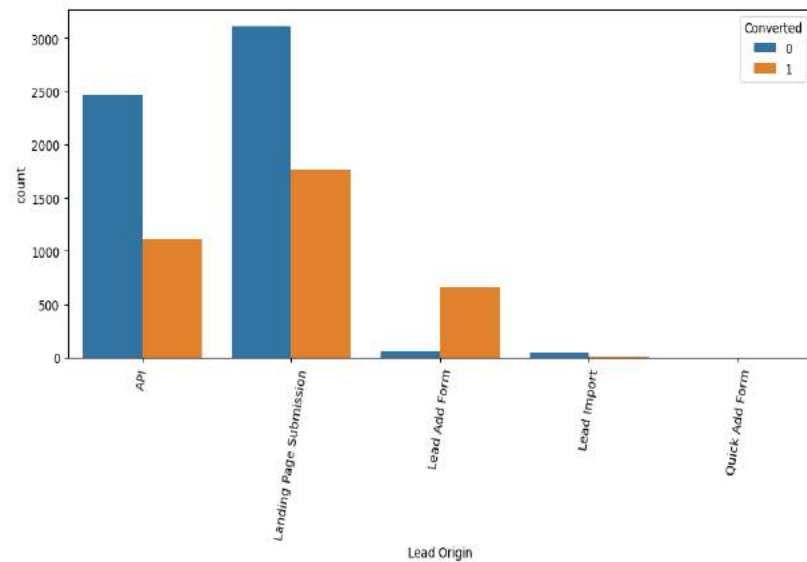
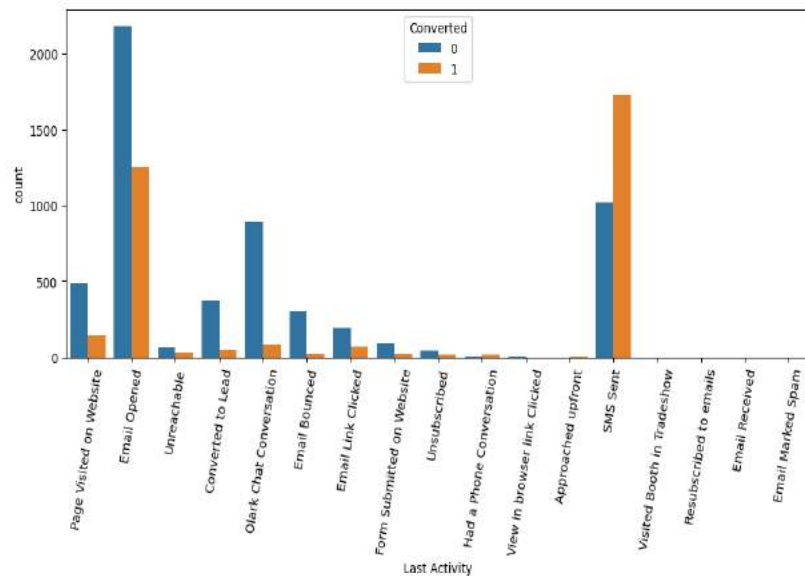
Data Cleaning and Data Manipulation

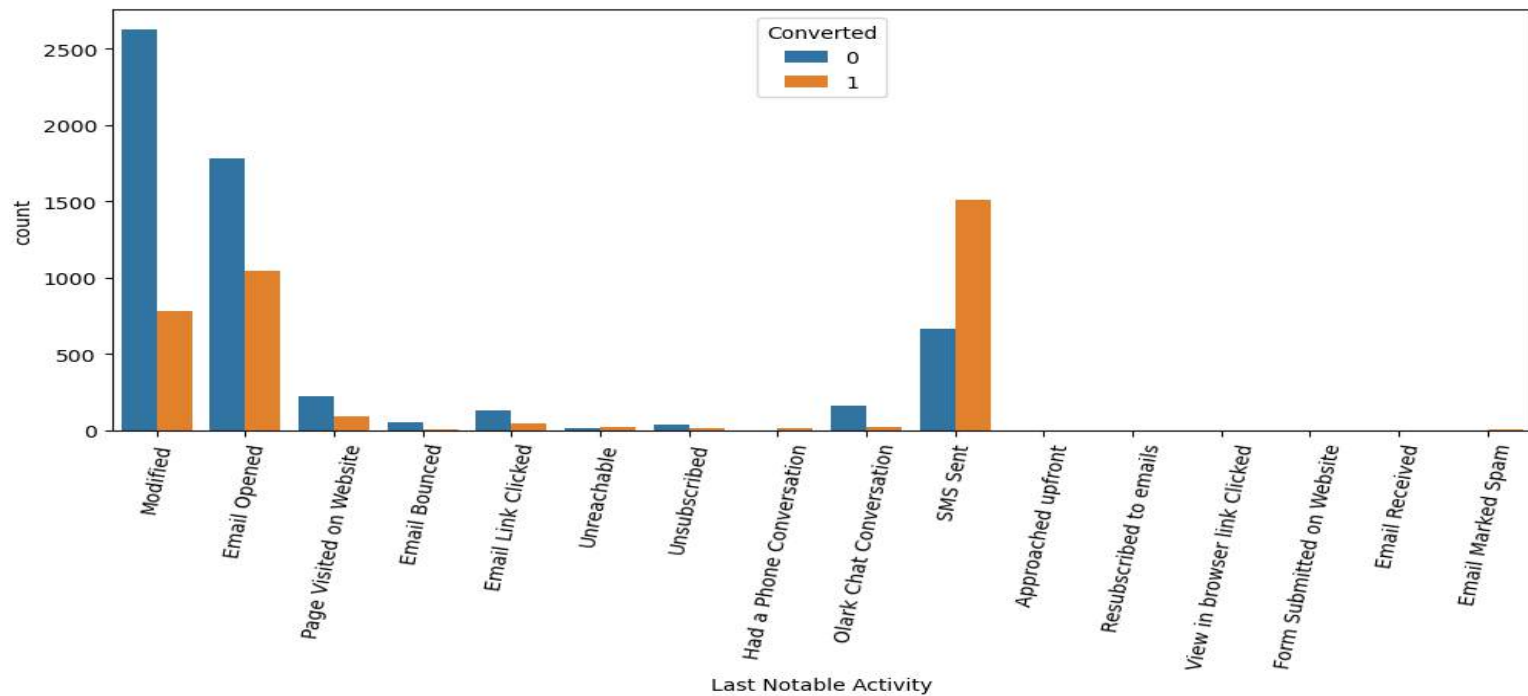
- Total number of Rows = 9240, Total number of Columns=37.
- Removing 'Lead Number' and 'Prospect ID' which is not necessary for analysis since they all have Unique values.
- Converting 'Select' values to 'NAN'.
- Dropping unique valued Columns.
- Dropping off the Columns with more than 35% missing values.

EDA (Categorical Attributes Analysis)

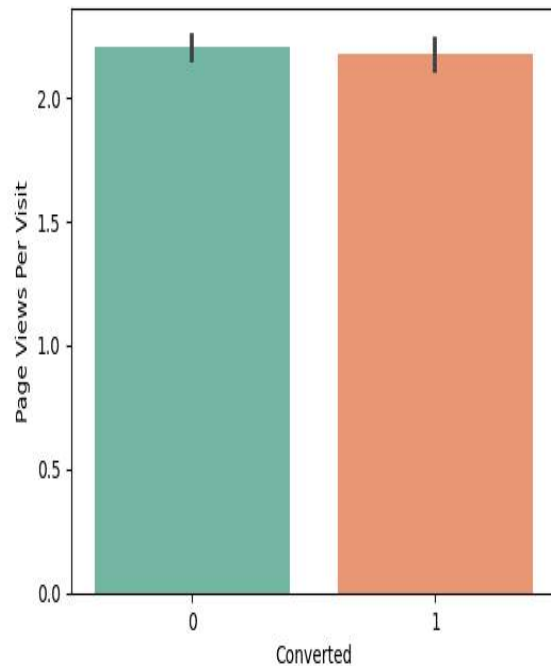
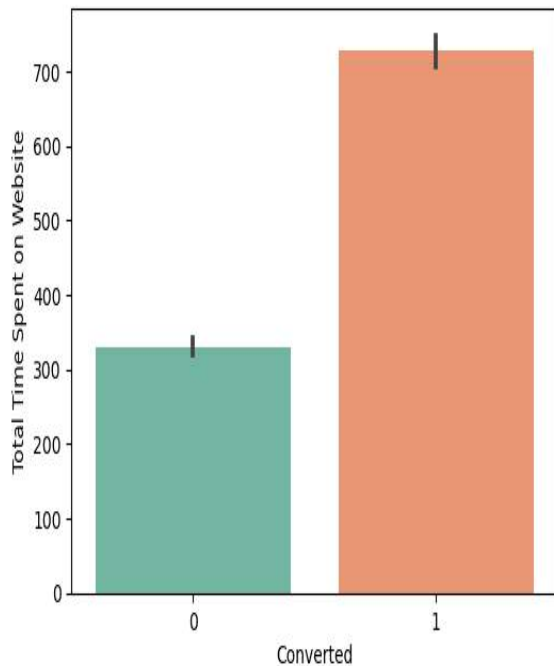
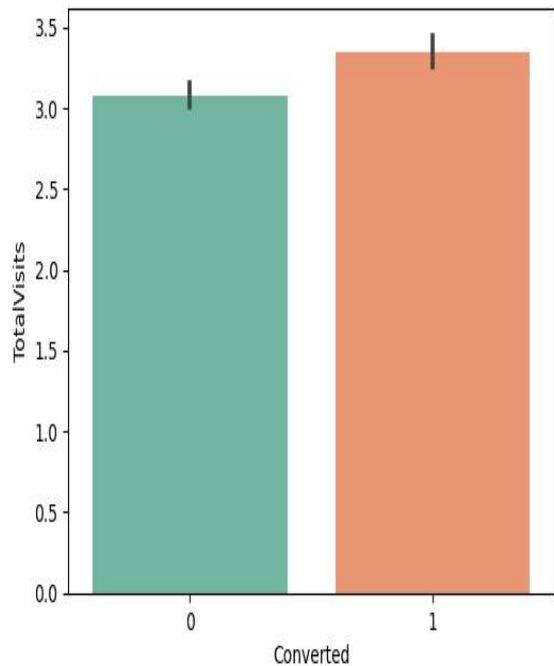








EDA (Numerical Variables)



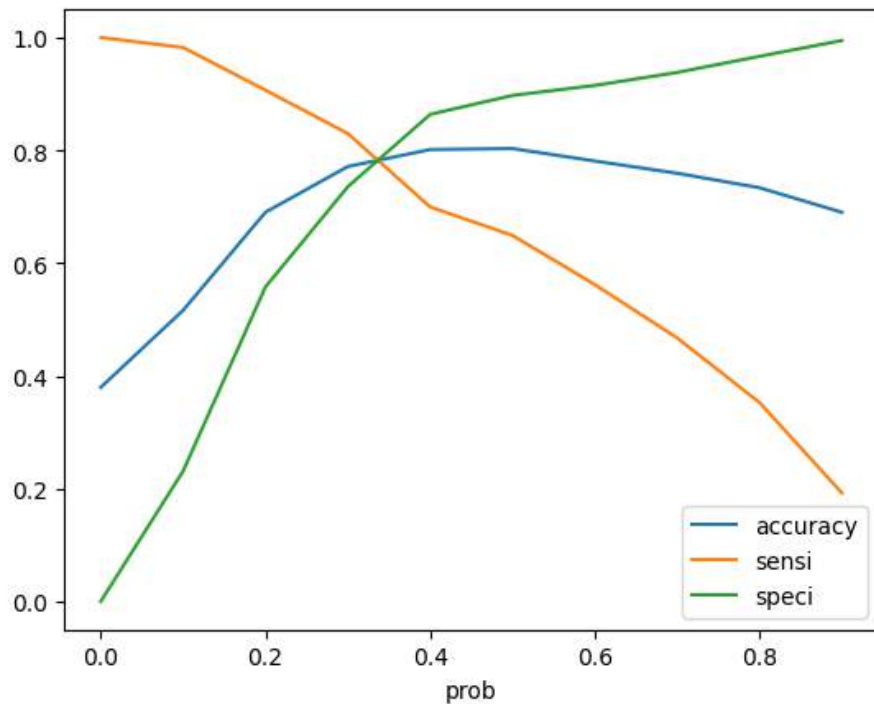
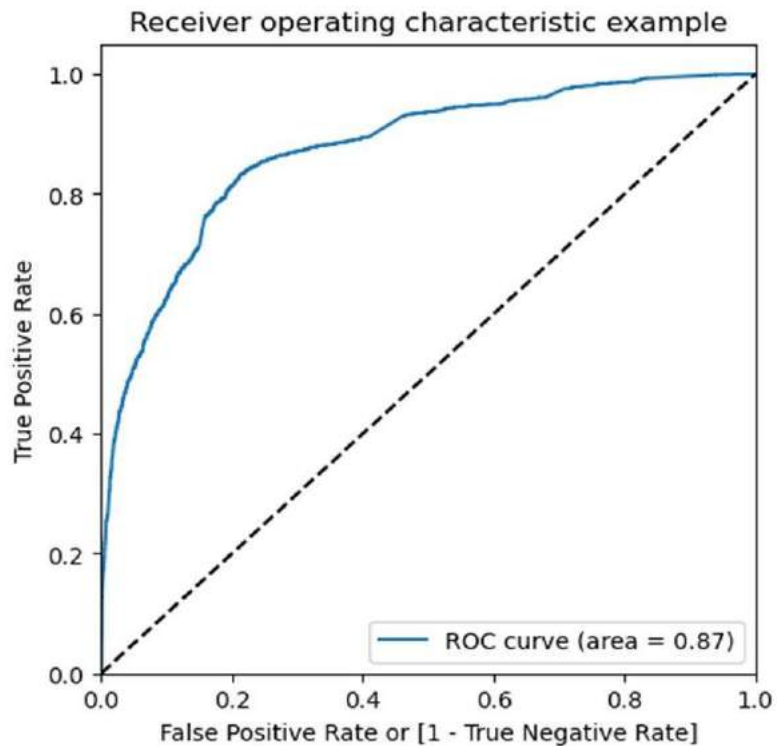
Data Conversion

- Numerical Variables are Normalised.
- Dummy Variables are created for Object Type Variables.
- Total Rows for Analysis: 9103
- Total Columns for Analysis: 23

Model Results

- Splitting the Data into Training and Testing Sets.
- The first basic step for Regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection.
- Running RFE with 15 variables as output.
- Building Model by removing the variable whose p-value is greater than 0.05 and VIF value is greater than 5.
- Predictions on Test Data Set.
- Overall accuracy is 80.3%.

ROC Curve



ROC Curve (Continued)

- Finding Optimal Cut Off Point.
- Optimal Cut Off Probability is that Probability where we get balanced Sensitivity and Specificity.
- From the second graph, it's visible that the optimal cut-off is approximately at 0.30.

Conclusion and Recommendations

- **Looking at the below variables, we should be able to help X-Education company with finding out hot leads which will be very high potential customers:-**
 1. Do Not Email
 2. Total Time Spent on Website
 3. Lead Origin_Lead Add Form
 4. Lead Source_Direct Traffic
 5. Lead Source_Google
 6. Lead Source_Organic Search
 7. Lead Source_Reference
 8. Lead Source_Referral Sites
 9. What is your current occupation_Other
 10. What is your current occupation_Student
 11. What is your current occupation_Unemployed
 12. What is your current occupation_Working Professional
- **Taking the above details into consideration and using this model, the X-Education company can maximise the sales of their course and also generate an excellent revenue.**

THANK YOU!!!