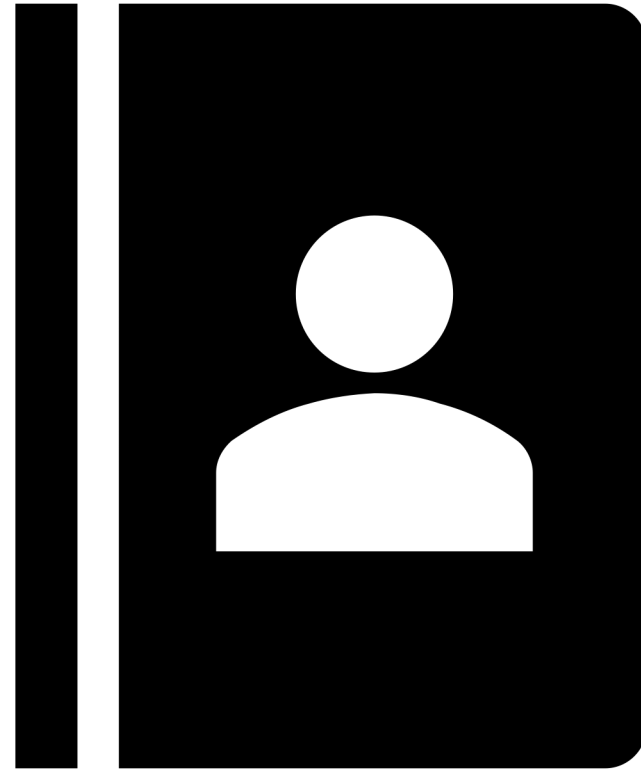

LEAD SCORING CASE STUDY

By

- 1) Godwin Geo Rai
- 2) Gayatri Bharatwaj
- 3) Satyanarayana Murthy



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

1) Data cleaning and data manipulation

- Check and handle duplicate data.
- Check and handle NA values & missing values.
- Drop columns, if it contains large amount of missing values and not useful for the analysis.
- Imputation of the values ,If necessary.
- Check & handle outliers in data.

2) EDA

- Univariate data analysis : Value count ,distribution of variables etc.
 - Bivariate data analysis : Correlation, coefficients & pattern between the variables etc.
-

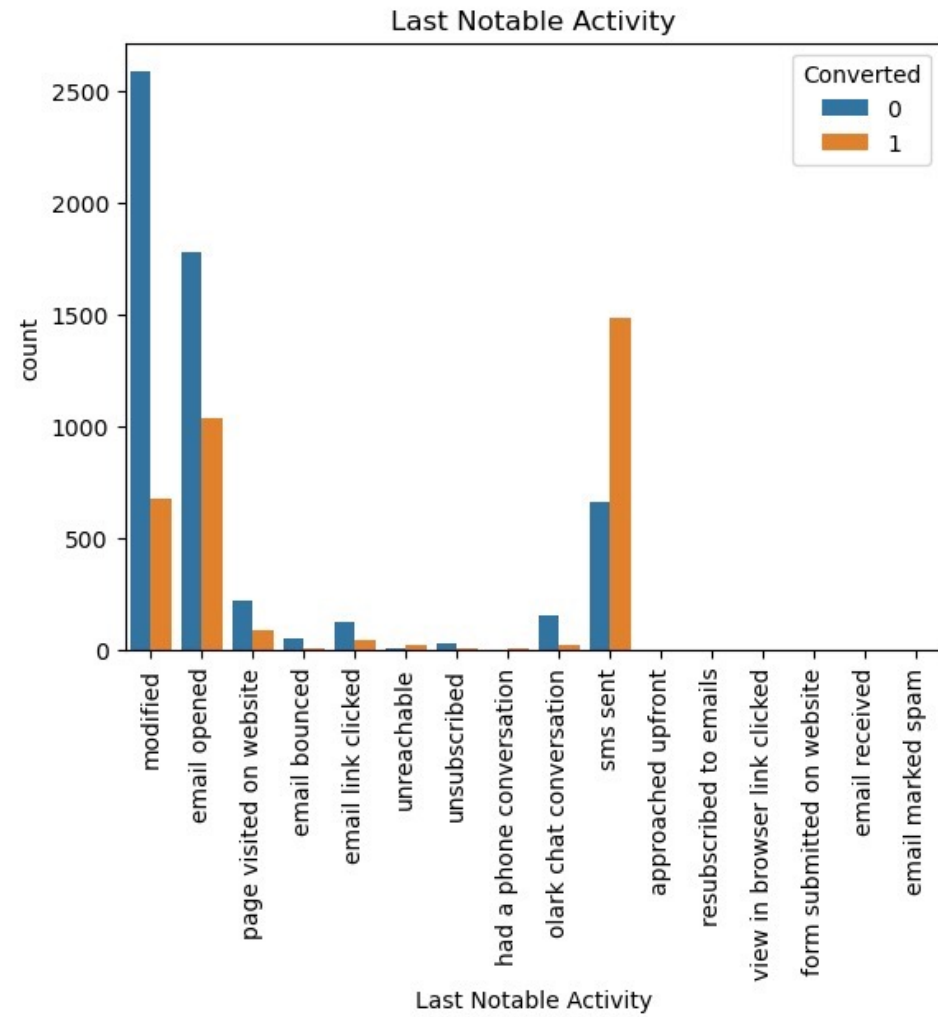
-
- Feature scaling & dummy variables and encoding of the data.
 - Classification technique : Logistic regression used for the model making & prediction.
 - Validation of the model.
 - Model presentation.
 - Conclusions & recommendations.

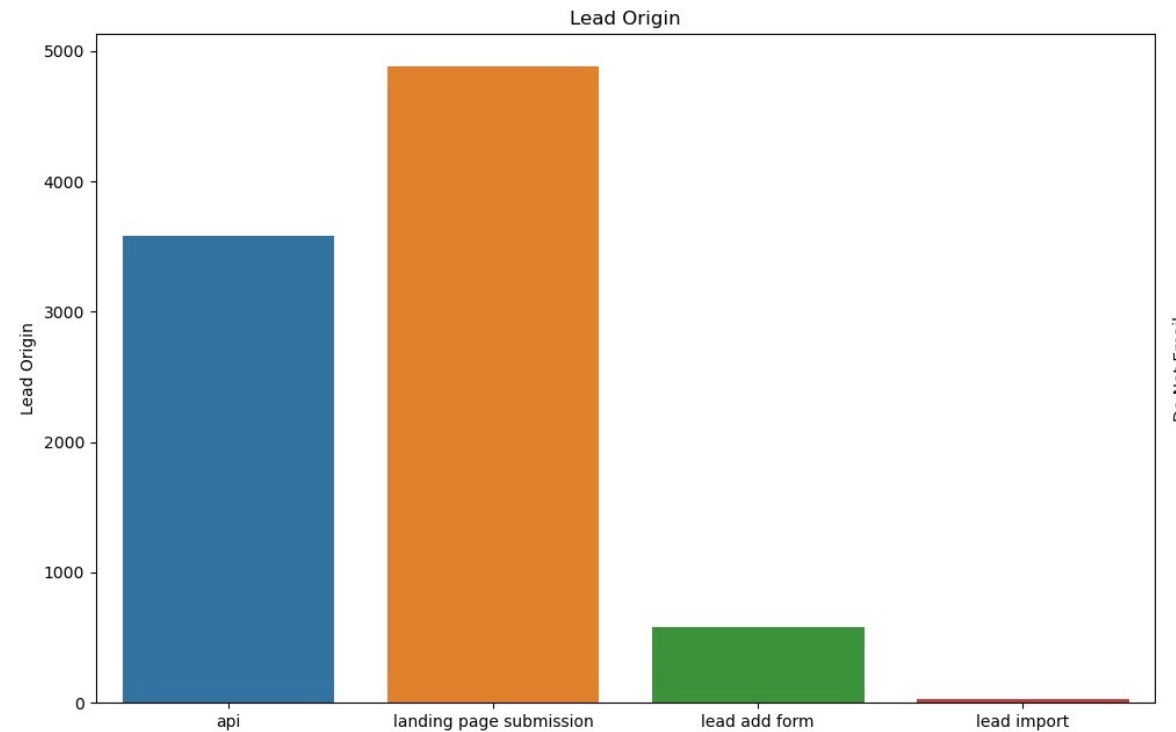
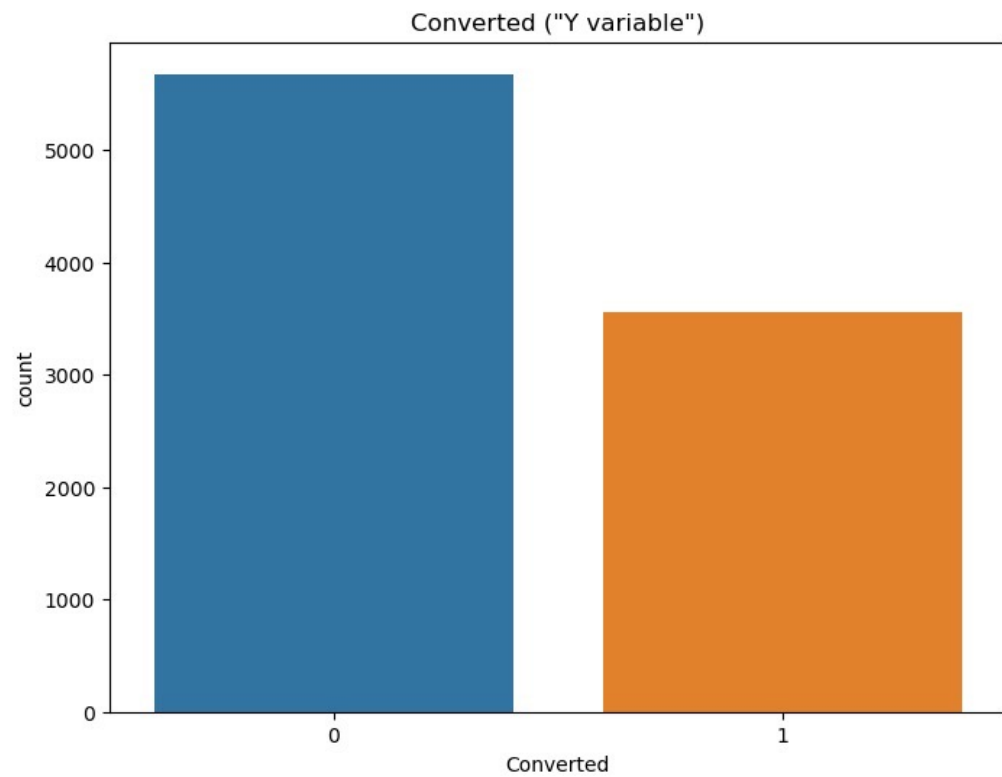
Data Manipulation

- Total Number of Rows =37, Total Number of Columns =9240.
 - Single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply” .
 - Chain Content”, “Get updates on DM Content”, “I agree to pay the amount through cheque” etc. have been dropped.
 - Removing the ‘Prospect ID’ & ‘Lead Number’ which is not necessary for the analysis.
-

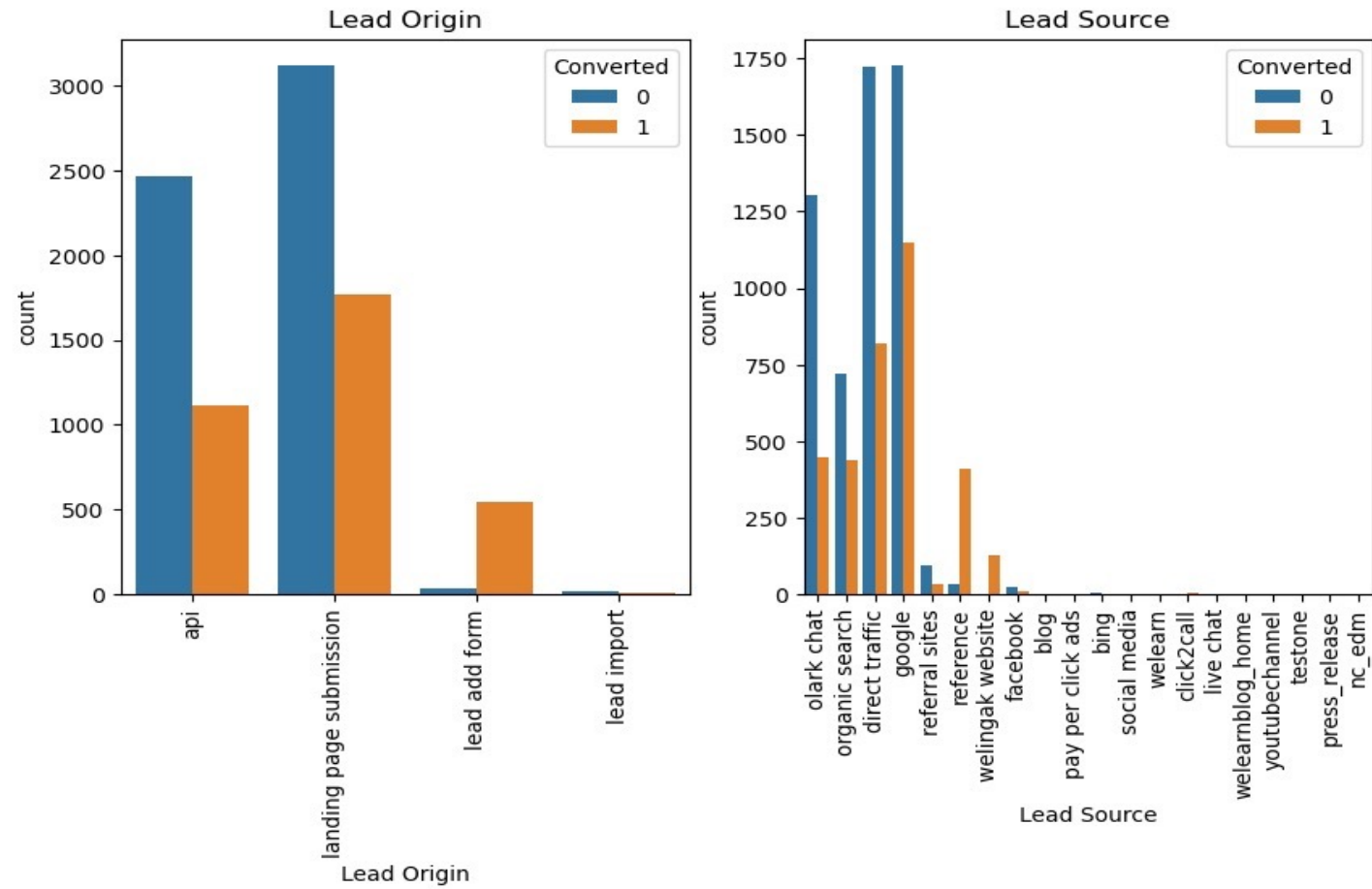
-
- After checking for the value counts for some of the object type variables, we find some of the features which has no enough variance, which we have dropped, the features are: “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.
 - Dropping the columns having more than 35% as missing value such as ‘How did you hear about X Education’ and ‘Lead Profile’.
-

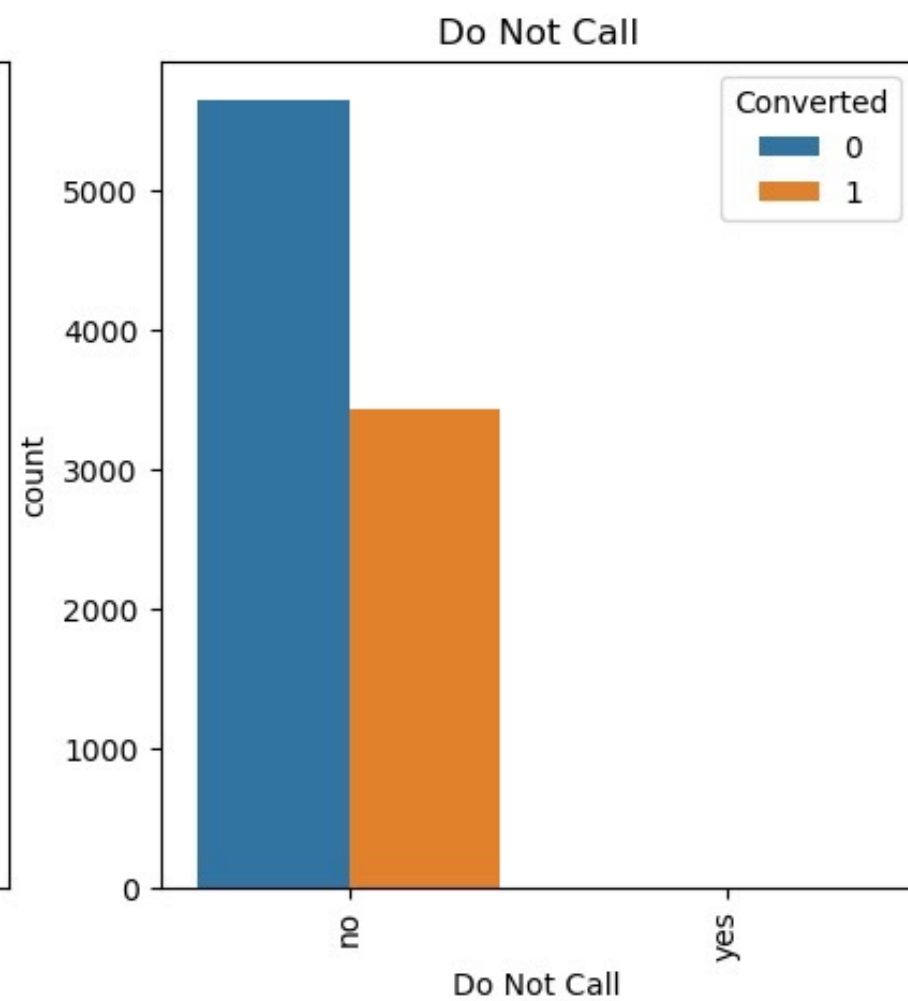
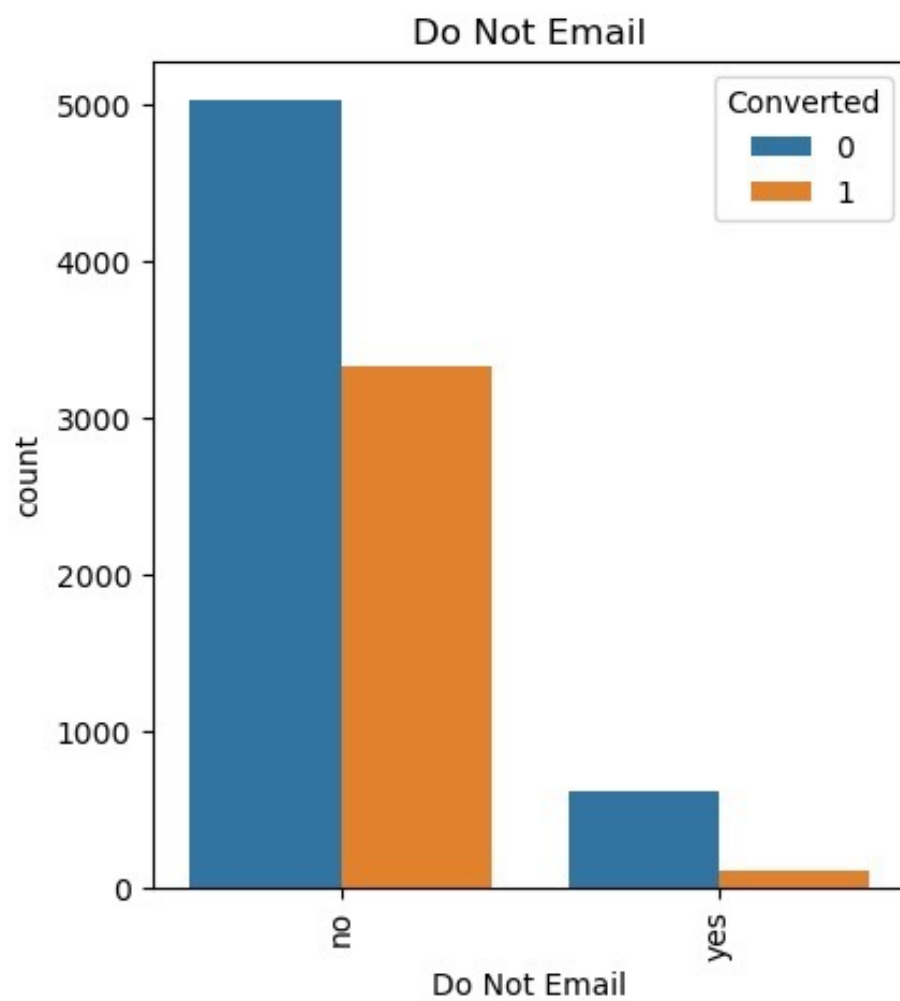
EDA

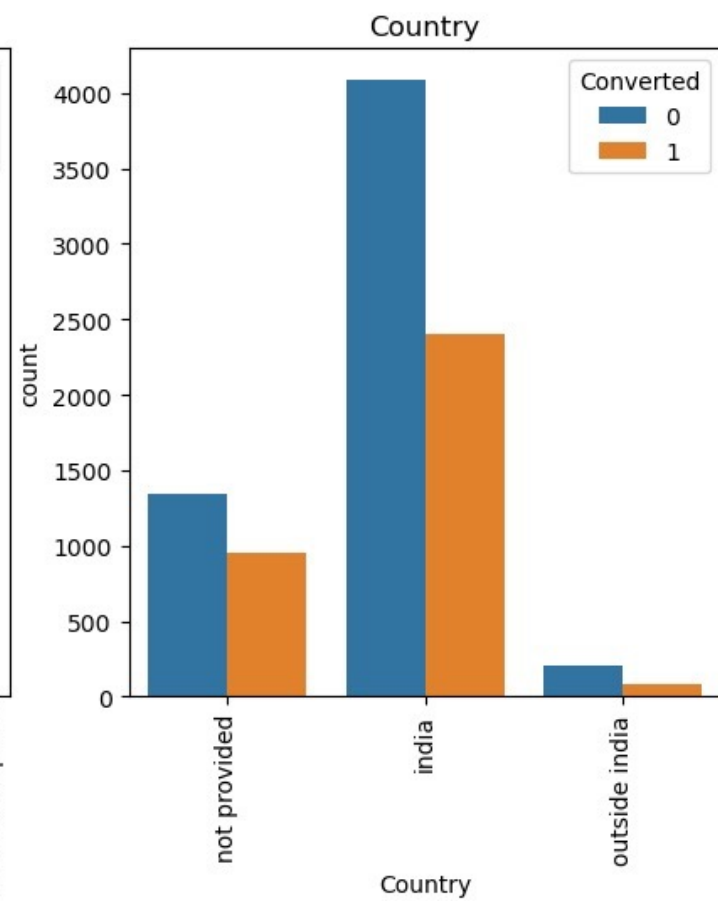
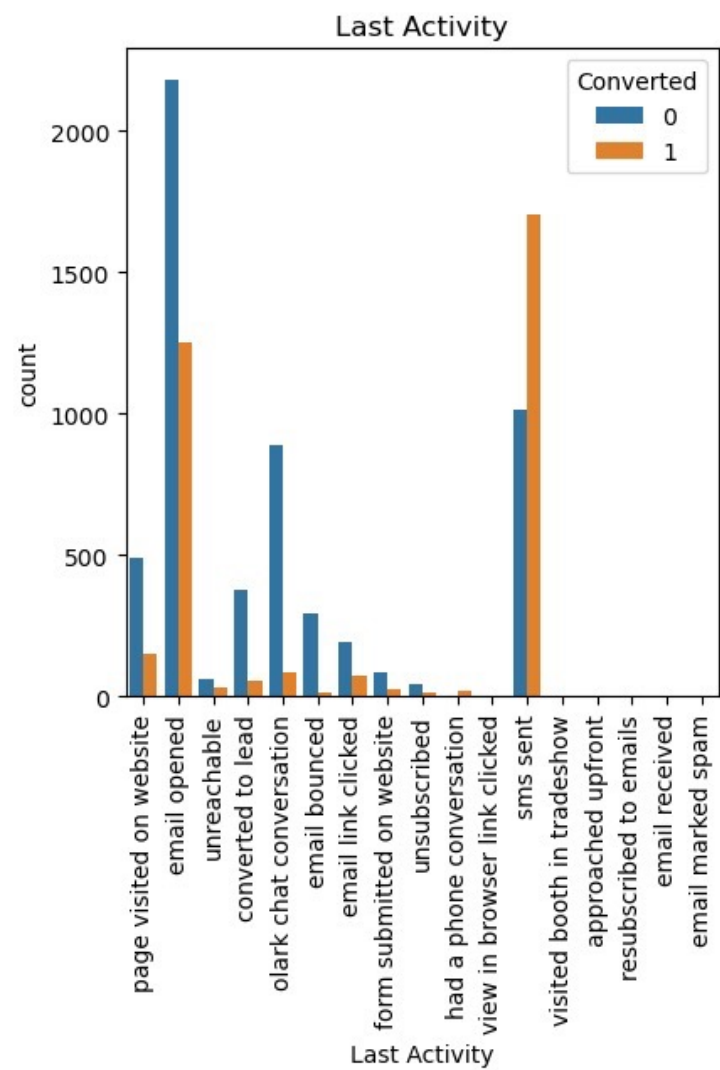




CATEGORICAL VARIABLE RELATION







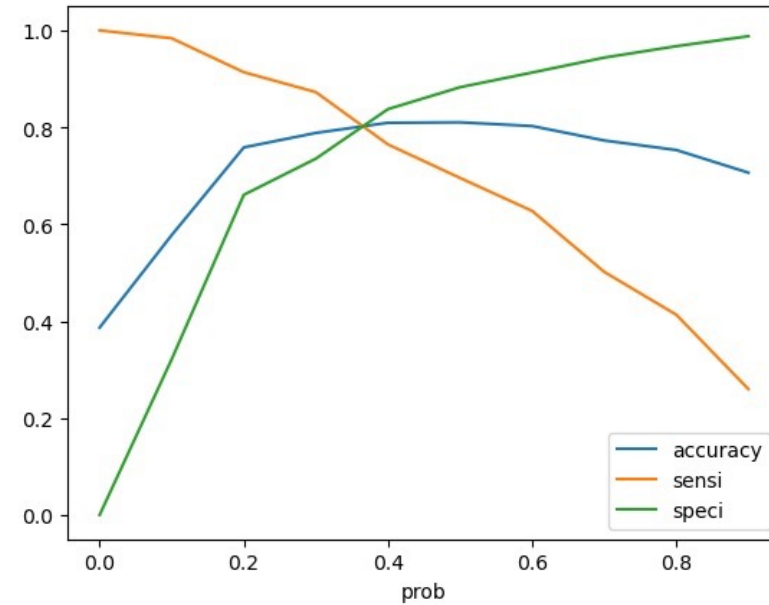
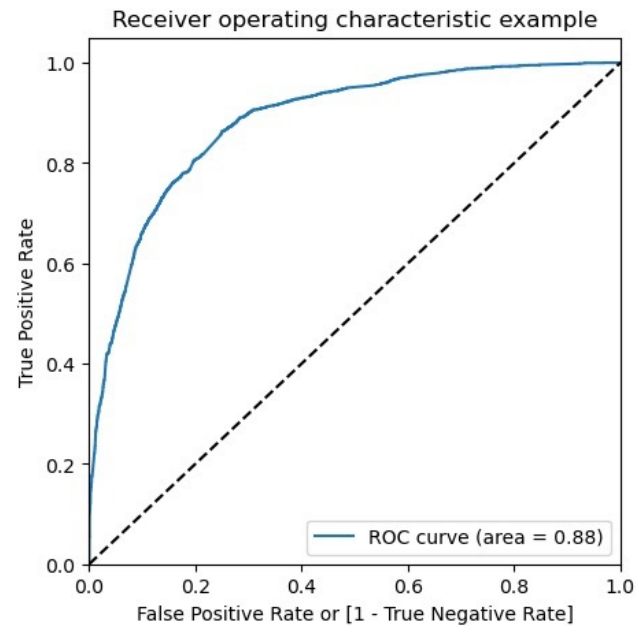
DATA CONVERSION

- Numerical Variables are normalized.
 - Dummy variables are created for object type variables.
 - Total rows for analysis : 8792
 - Total columns for analysis : 43
-

MODEL BUILDING

- Splitting the data into training & 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 variables whose p-value is greater than 0.05 & VIF value is greater than 5.
 - Predictions on test data set.
 - Overall accuracy 81%.
-

ROC CURVE



- Finding optimal cut off point is that the probability where we get balanced sensitivity & specificity.
 - From the second graph it is visible that the optimal cut off is at 0.35.
-

CONCLUSION

The most influential variables for potential buyers were identified, ranked as follows:

1. Total time spent on the website.
2. Total number of visits.
3. Lead sources, including Google, Direct traffic, Organic search, and Welingak website.
4. Last activity, particularly SMS and Olark chat conversation.
5. Lead origin as Lead add format.
6. Current occupation as a working professional.

Implications:

- X Education stands to benefit significantly by targeting potential buyers based on these key variables, thereby increasing the likelihood of course enrolment among industry professionals.
-