



Lead Scoring Case Study

GROUP MEMBERS:
SHIVANI SHRIVASTAVA
SATWIK 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.

Business Problem Solution

- ▶ **Data cleaning and data manipulation.**

- ▶ Check and handle duplicate data.
- ▶ Check and handle NA values and 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 and handle outliers in data.

- ▶ **EDA**

- ▶ Univariate data analysis: value count, distribution of variable etc.
- ▶ Bivariate data analysis: correlation coefficients and pattern between the variables etc.

Business Problem Solution

Continued...

- ▶ Feature Scaling for numerical columns
- ▶ Dummy Variables creation and encoding of the data for categorical columns.
- ▶ Classification technique: logistic regression used for the model making and prediction.
- ▶ Validation of the model.
- ▶ Model presentation.
- ▶ Conclusions and recommendations.

Business Problem Solution

Continued...

► Data Manipulation

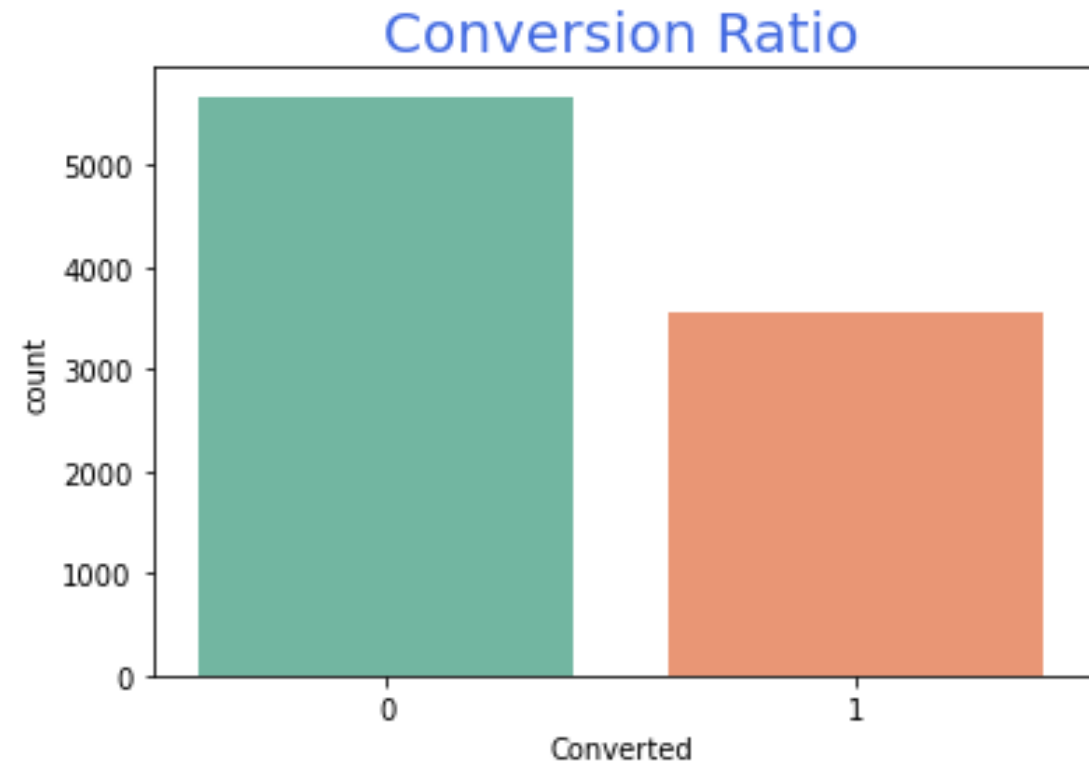
1. Total Number of Rows =37, Total Number of Columns =9240.
2. Single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply”
3. Chain Content”, “Get updates on DM Content”, “I agree to pay the amount through cheque” etc. have been dropped.
4. Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis.
5. 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:
6. “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.
7. Dropping the columns having more than 35% as missing value such as ‘How did you hear about X Education’ and ‘Lead Profile’.

Business Problem Solution

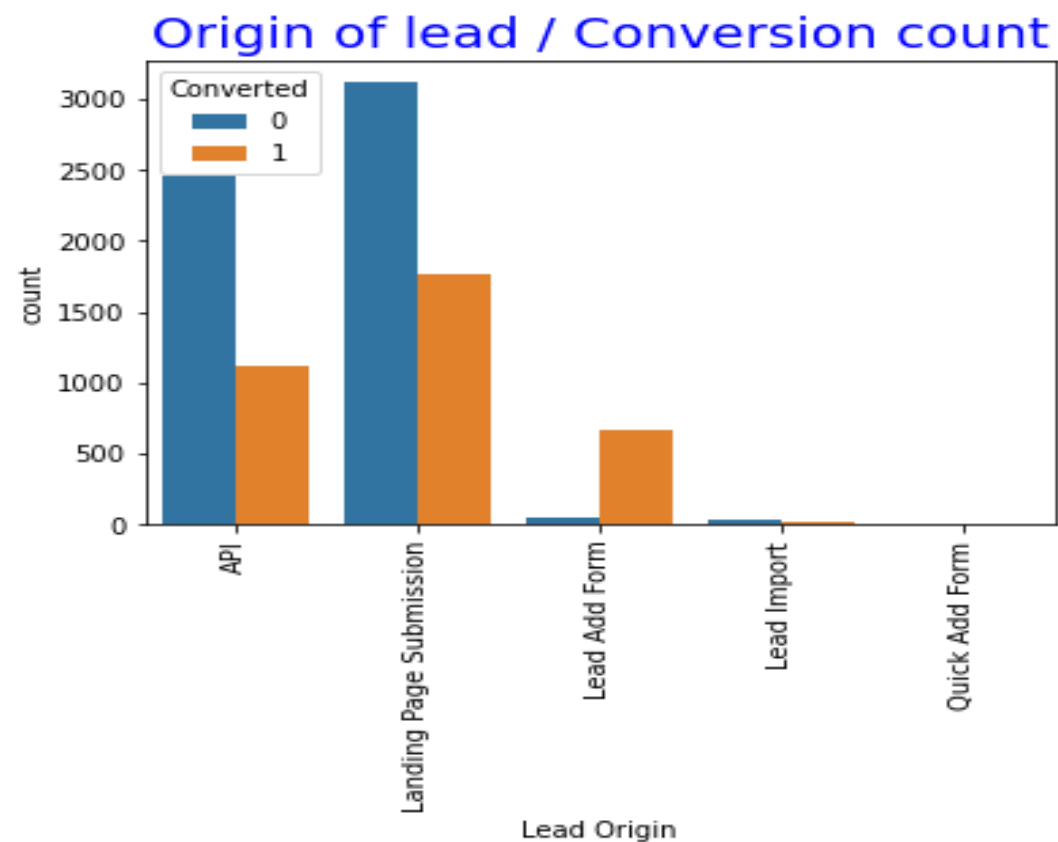
Continued...

► Exploratory Data Analysis

It can be observed that the conversion rate is around 39%.

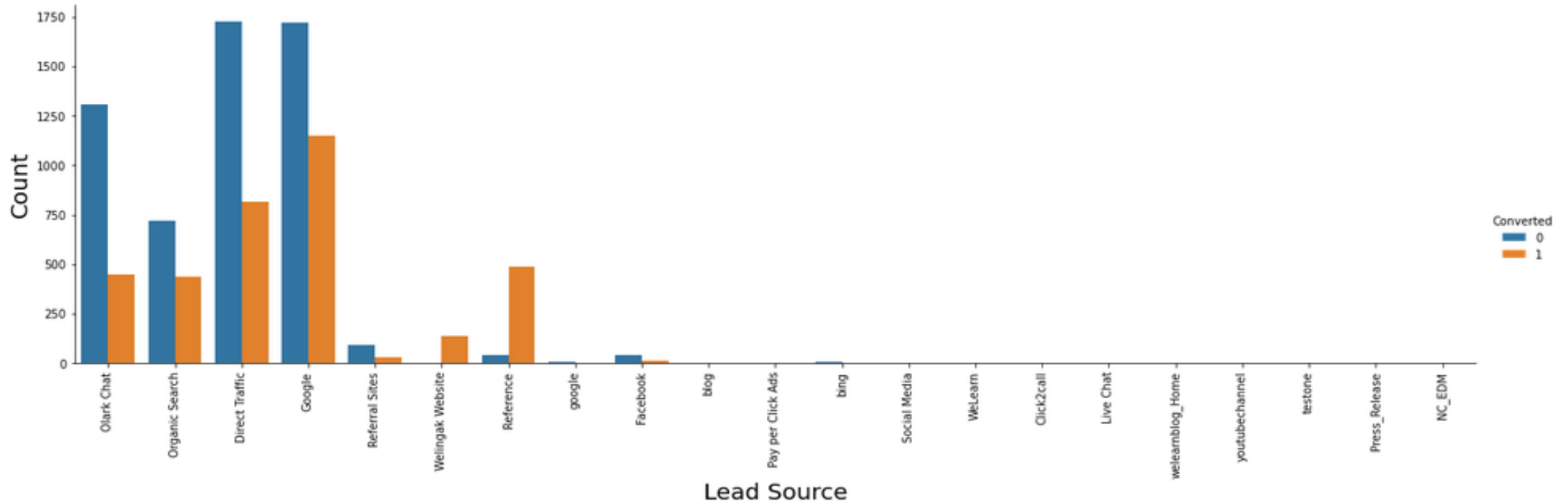


The highest amount of conversion happened by Landing Page Submission second to which is API. Only 1 was converted through Quick Add form.

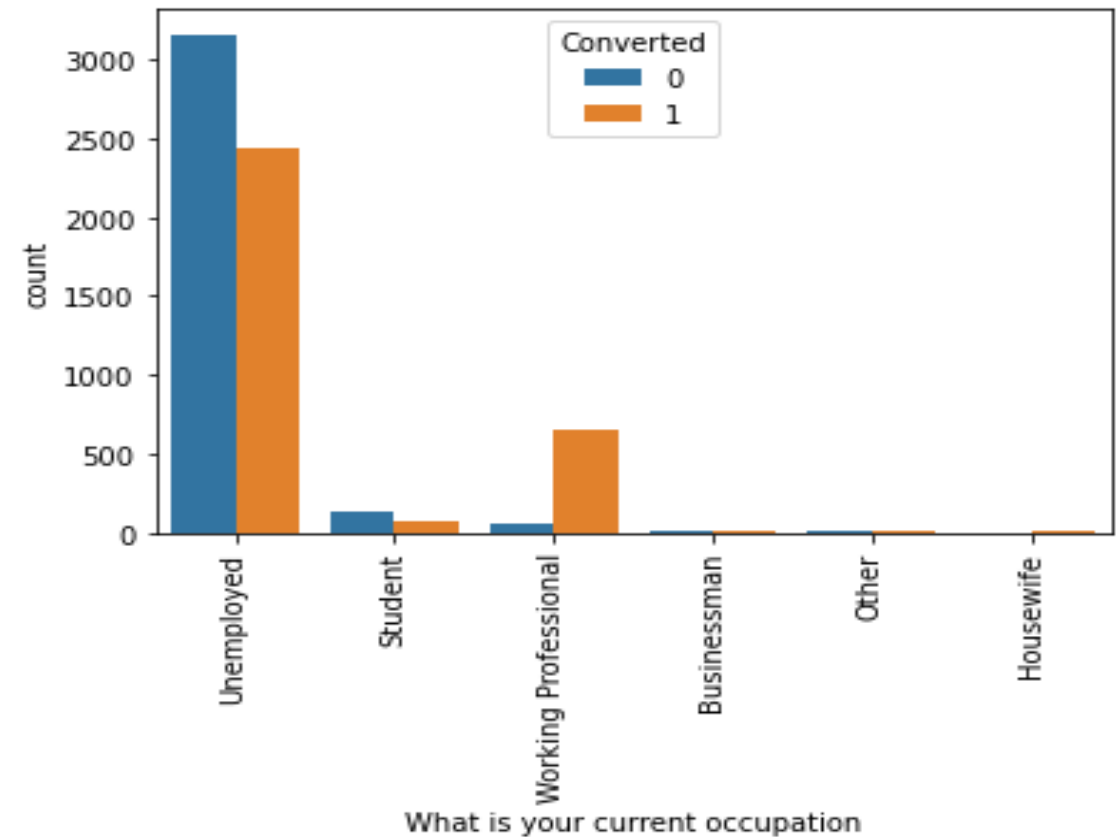


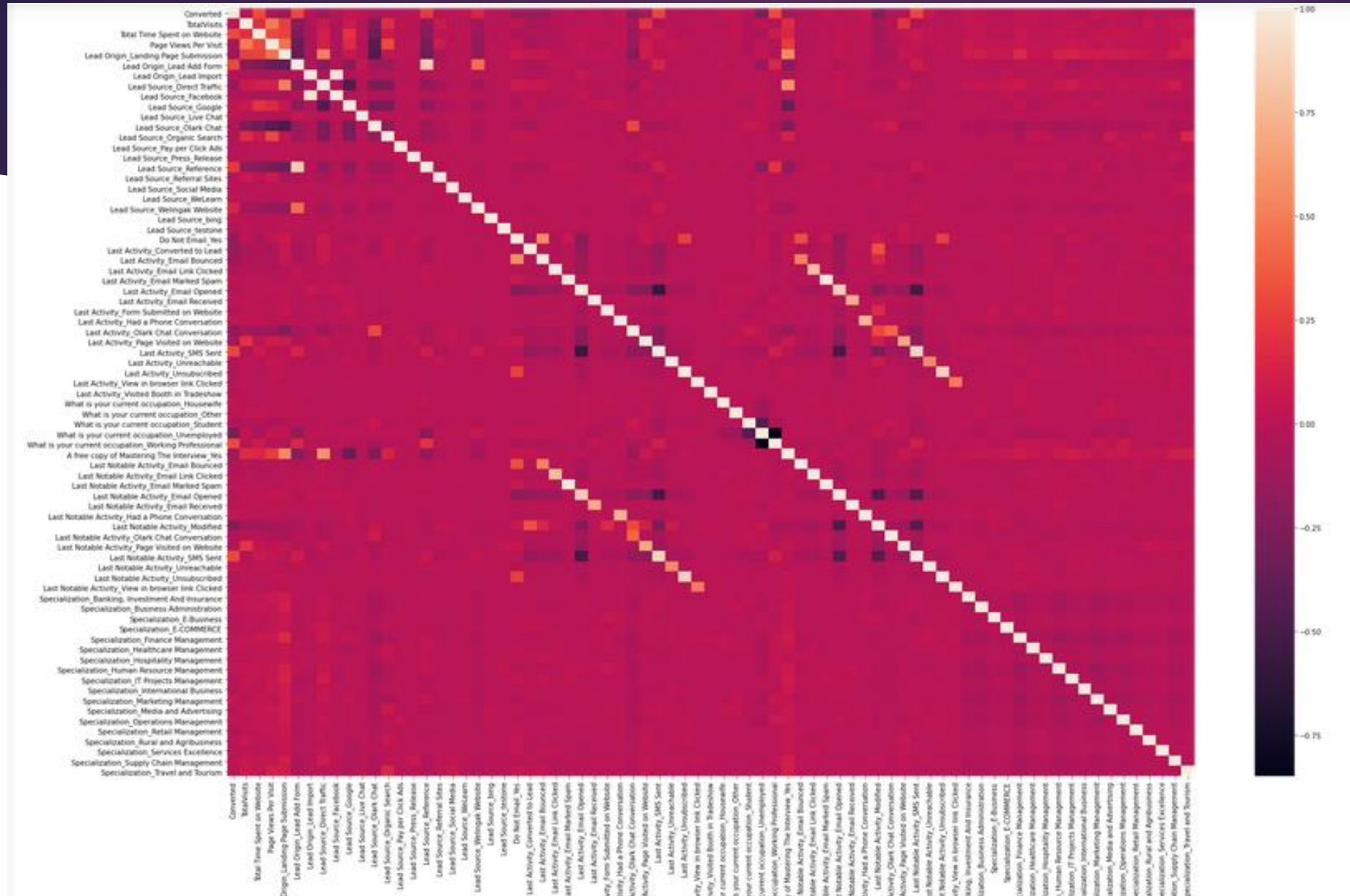
Although, it's known the lead's source to find a course/website is mostly through Google, This highest conversion rate is also from this group. Direct Traffic and Olark Chat following this in the respective order.

Lead Source Vs Converted



43 % Conversion rate for unemployed people
10 housewives were chosen as leads and all of them got converted.
More than 50% Conversion rate for students.
Since the data is not much in the 2nd and 3rd points. More data could be added for these for these to be validated.





No clear inferences could be drawn from heatmap

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

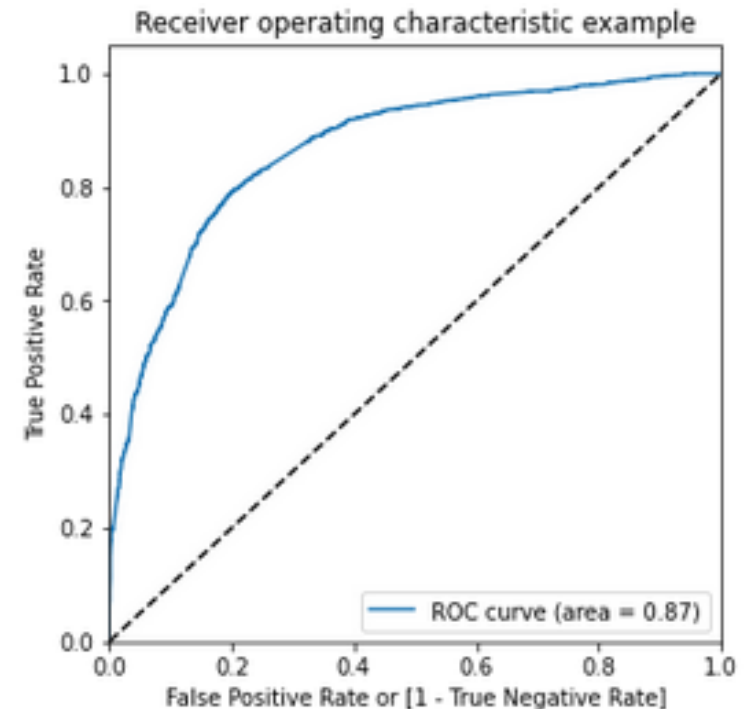
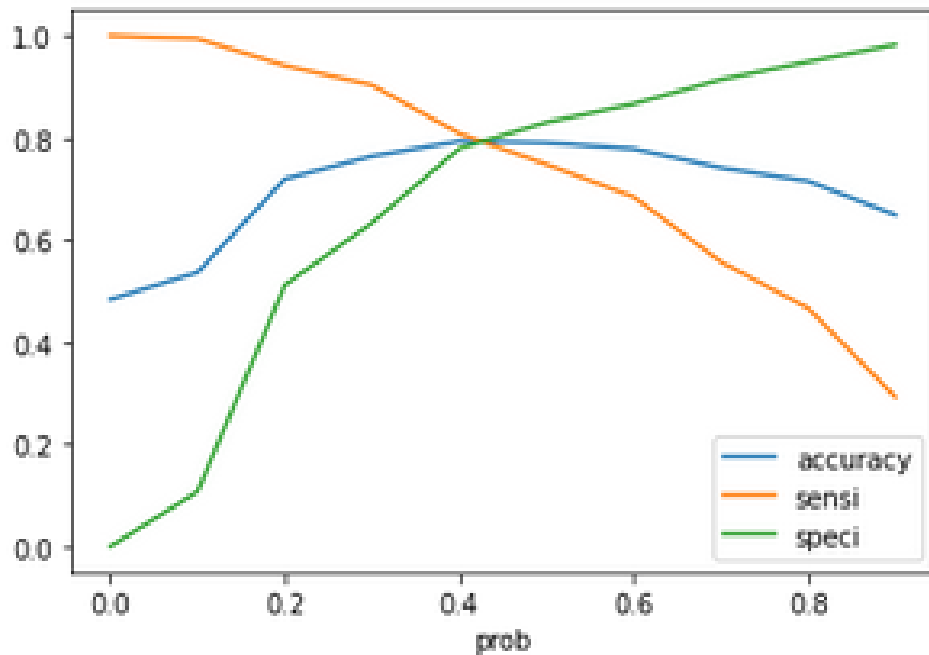
- ▶ Splitting the Data into Training and Testing Sets
- ▶ The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- ▶ Use RFE for Feature Selection
- ▶ Running RFE with 15 variables as output
- ▶ Building Model by removing the variable whose p- value is greater than 0.05 and VIF value is greater than 5
- ▶ Predictions on test data set
- ▶ Overall accuracy 80%

ROC Curve

(Finding Optimal Cut off Point)

The area under the curve is 0.87 which seems good. Now, we'll check the sensitivity-specificity trade-off to find the optimal cut-off.

It's observable that the optimal cut-off is around 0.43-0.42. We'll use 0.42.



Precision-Recall View

- ▶ Precision for Train data: ~78%
- ▶ Recall for Train data: ~79%
- ▶ Precision for Test data: ~78%
- ▶ Recall for Test data: ~77%
- ▶ From above we can see that both Train and Test dataset have almost same Precision and Recall.

Conclusion

- ▶ It was found that the variables that mattered the most in the potential buyers are (In descending order) :
- ▶ The total time spend on the Website.
- ▶ Total number of visits.
- ▶ When the lead source was: a. Google b. Direct traffic c. Organic search d. Welingak website
- ▶ When the last activity was: a. SMS b. Olark chat conversation
- ▶ When the lead origin is Lead add format.
- ▶ When their current occupation is as a working professional. Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

Continued...

- ▶ There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.
- ▶ First, sort out the best prospects from the leads you have generated. 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted. Then, You must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies.
- ▶ Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads.
- ▶ A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects. Focus on converted leads.
- ▶ Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.