

LEAD SCORE CASE STUDY

Submitted by

1. Desai Nagraj
2. Chidige Divija
3. Franco F



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.

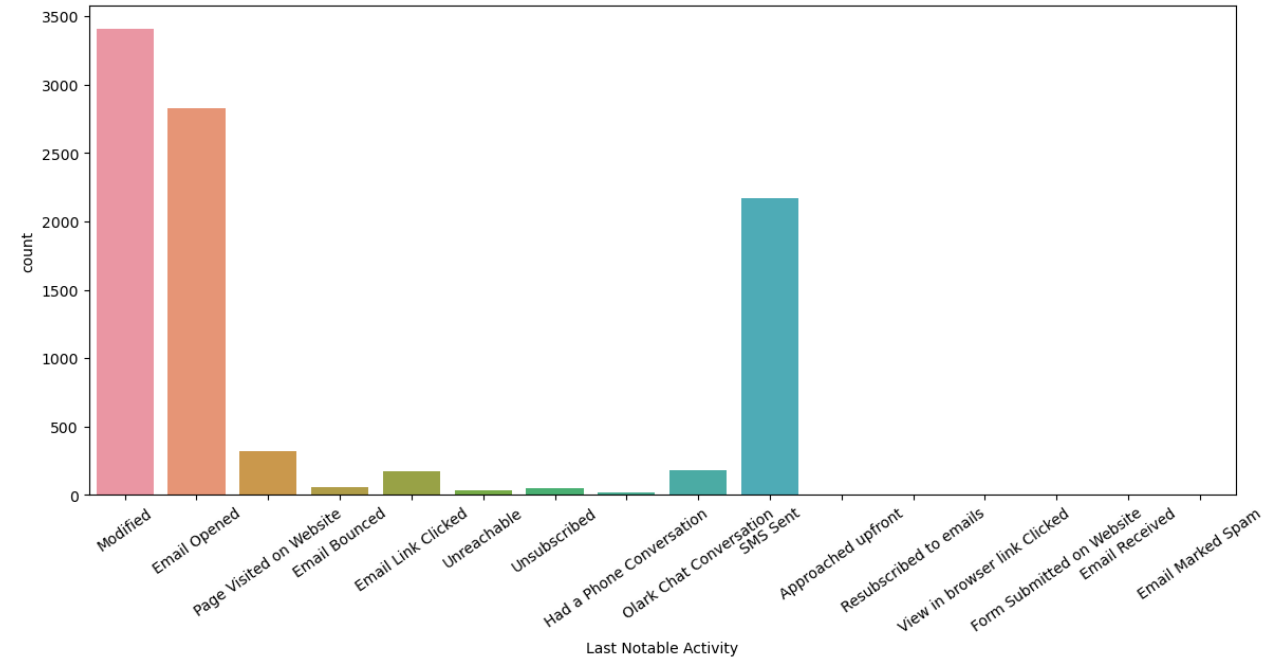
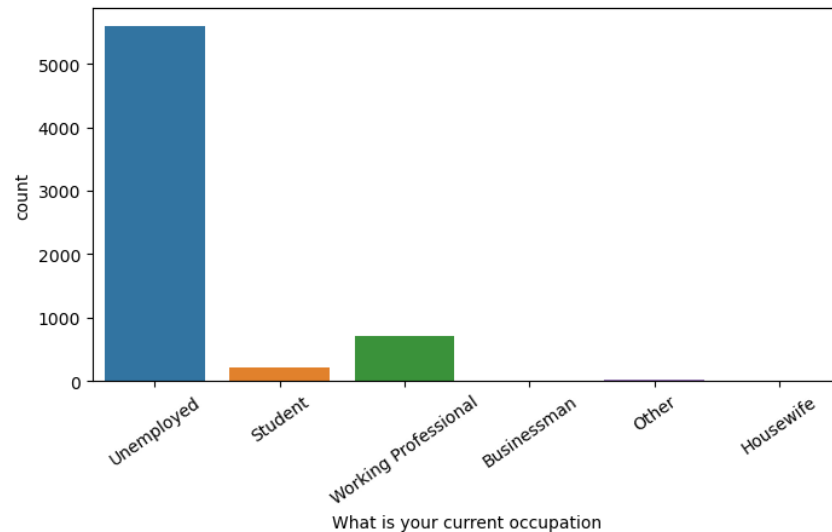
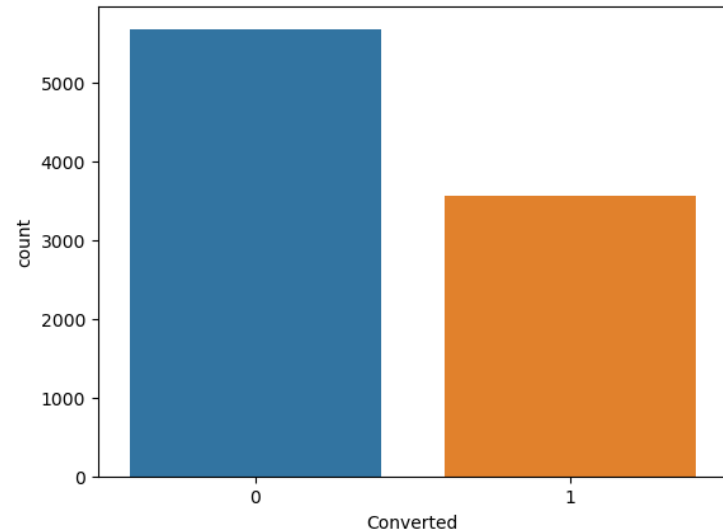


SOLUTION METHODOLOGY

- **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.
- Feature Scaling & Dummy Variables and encoding of the data.
- Classification technique: logistic regression used for the model making and prediction.
- Validation of the model.
- Model presentation.
- Conclusions and recommendations.



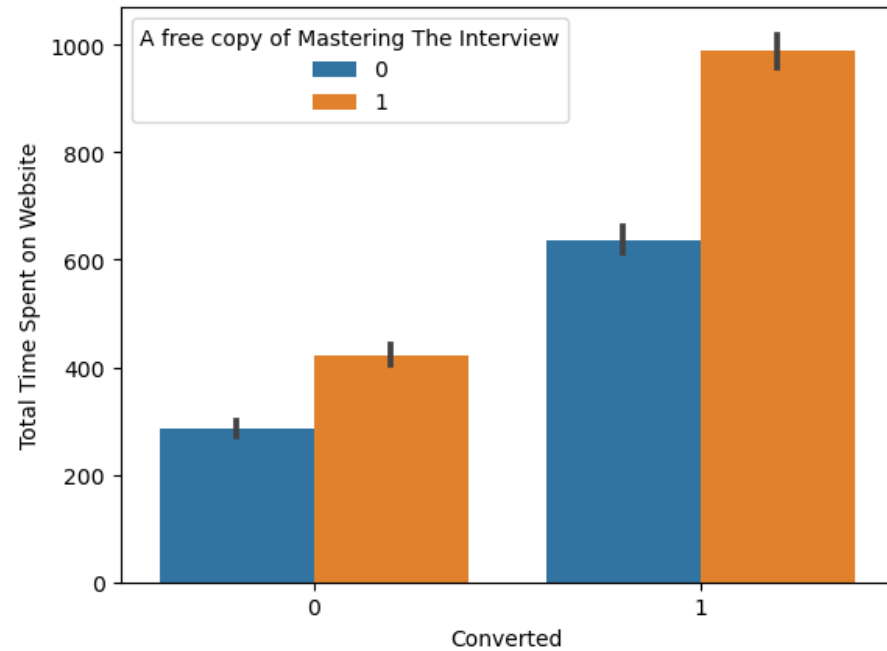
UNIVARIATE ANALYSIS



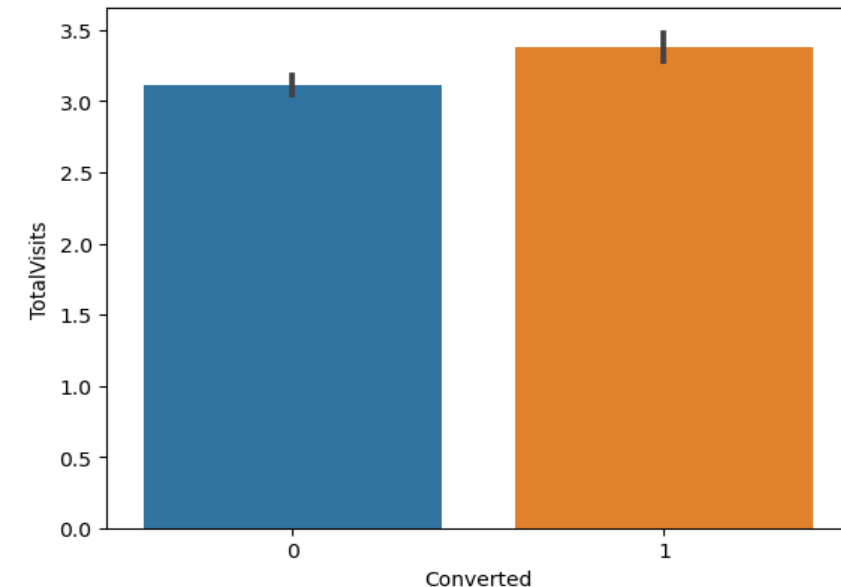
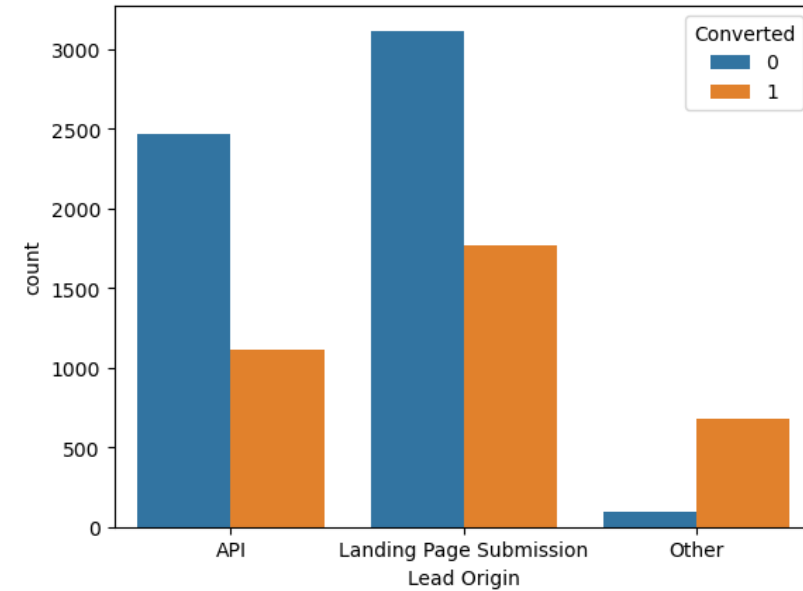
- Around 3k+ customers are converted
- Most of the customers are unemployed
- The last activity of most of the customer is modified, email opened and sms sent.



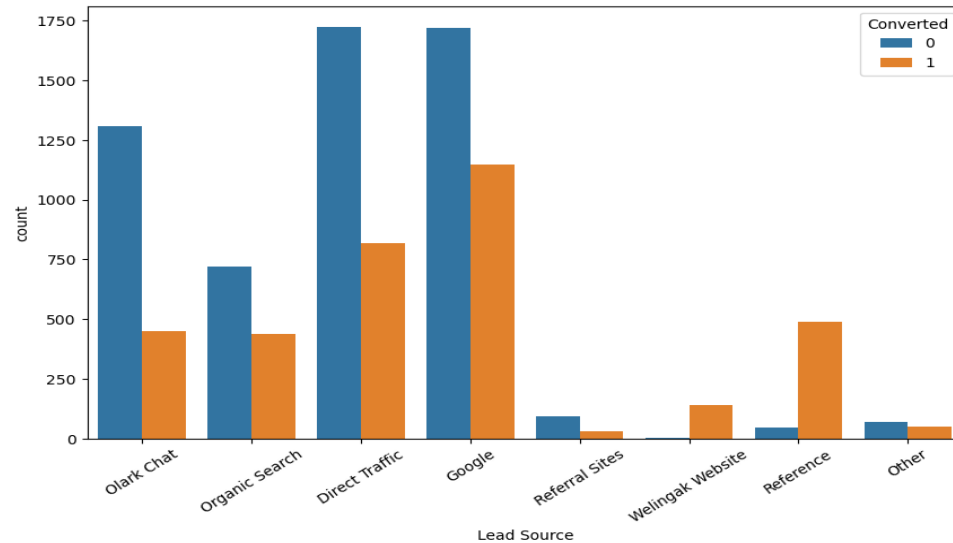
BIVARIATE ANALYSIS



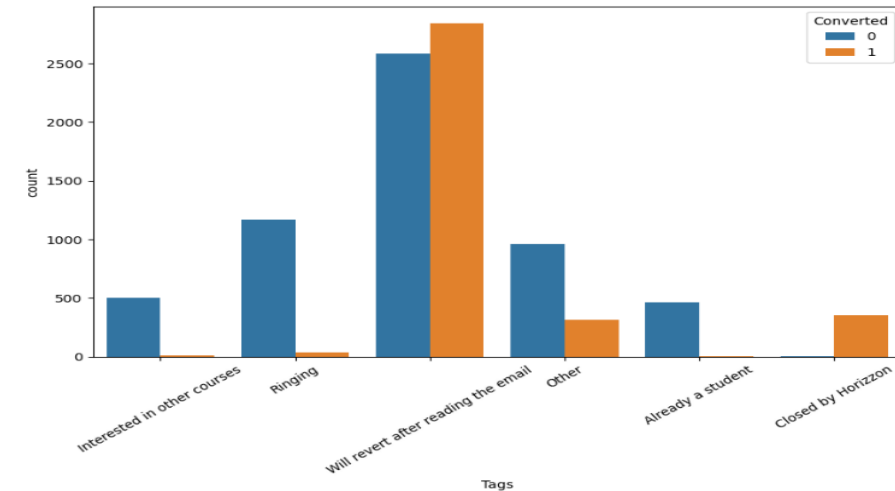
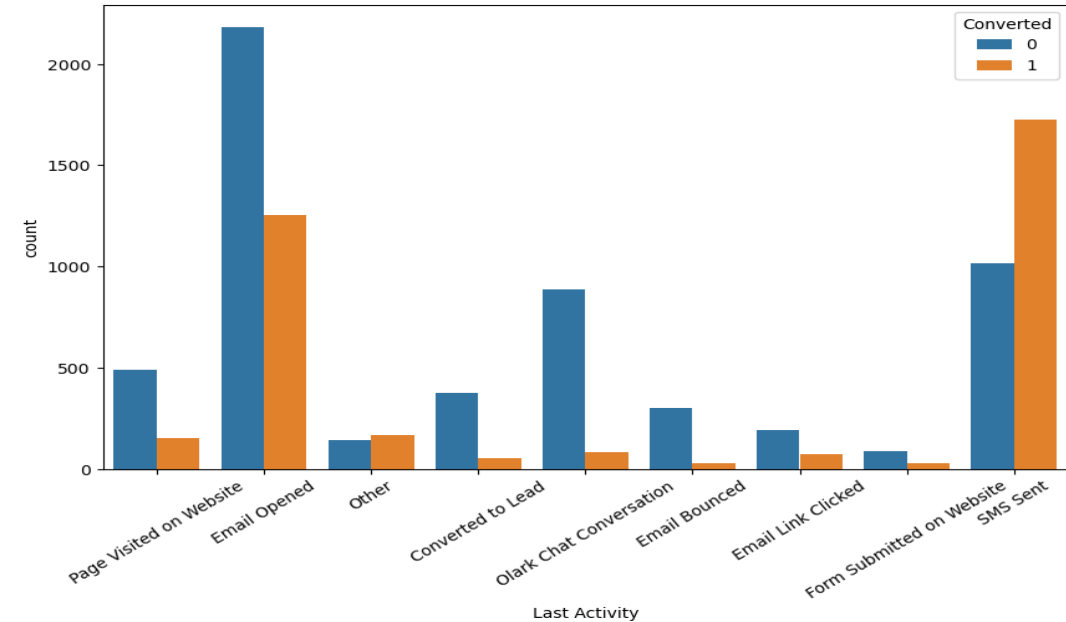
- Customers who are converted they spend more time on website
- Most of the customers lead origin is Landing Page Submission
- Customers who is converted are seen to visit more on website.



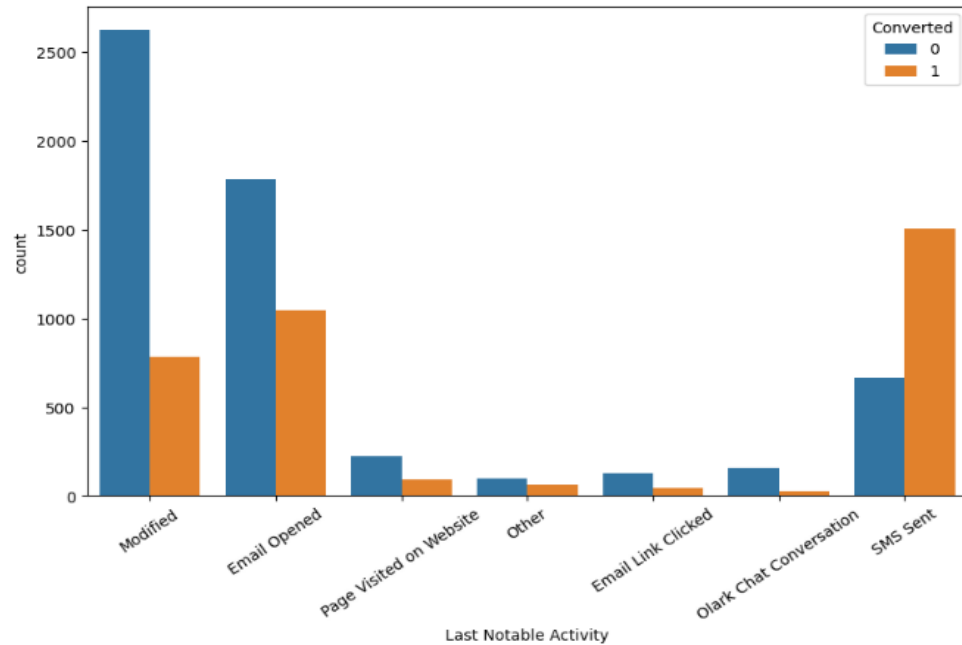
BIVARIATE ANALYSIS



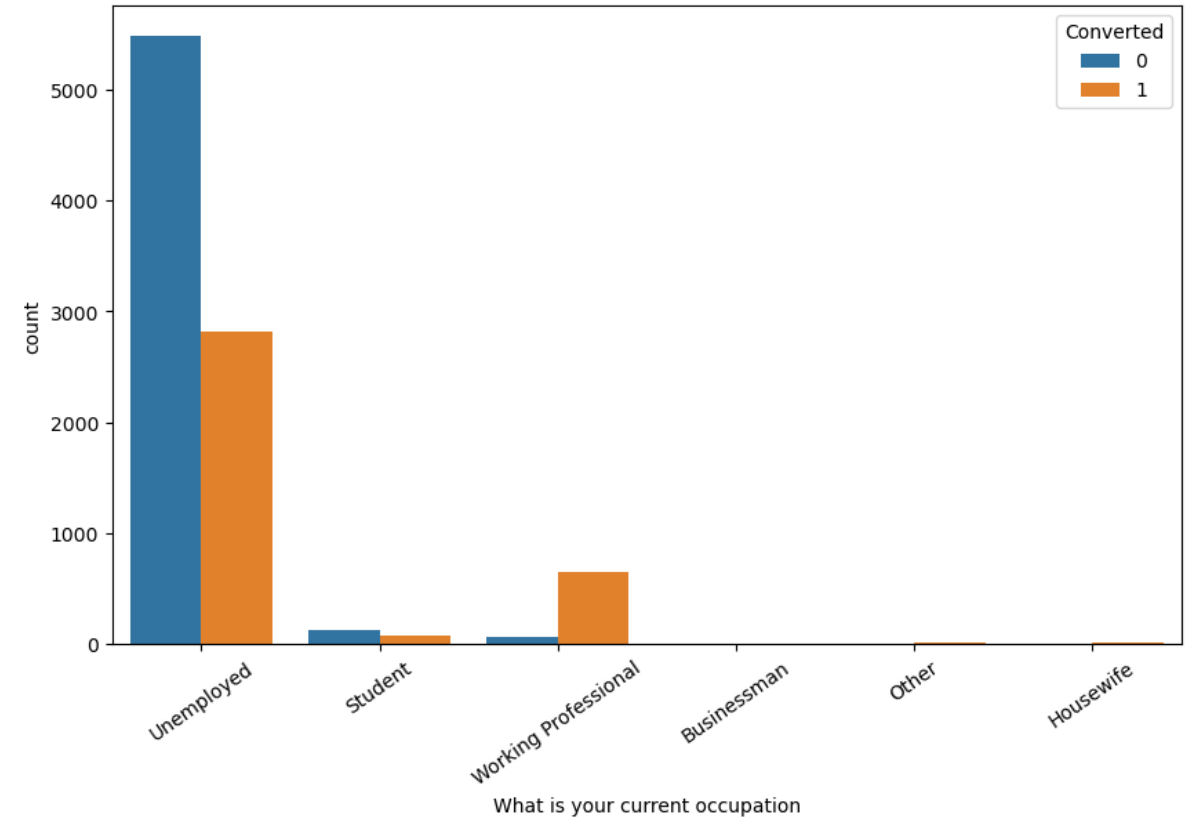
- Most of the customers lead source is Google
- Customers who come from reference maximum of them are converted
- Last activity of most of customers are Email opened and SMS sent.
- Customers with tags will revert after reading the email and Tags Closed by Horizzon has higher chances of conversion



BIVARIATE ANALYSIS



- Customers whose last notable activity is SMS sent has high chances of getting converted
- Customers who are working professional has more chances of getting converted

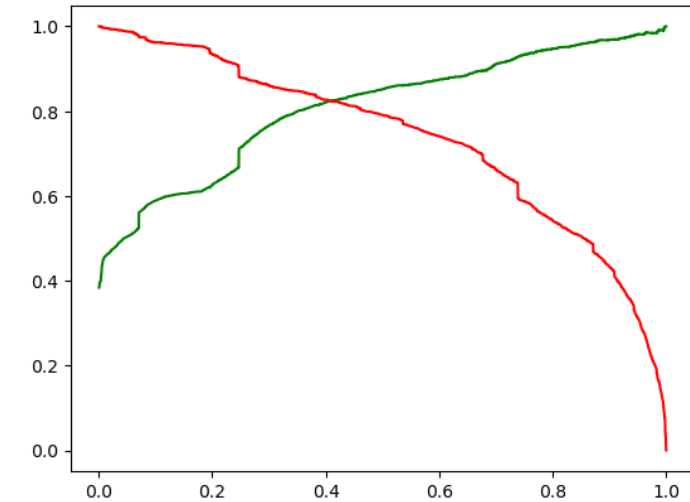
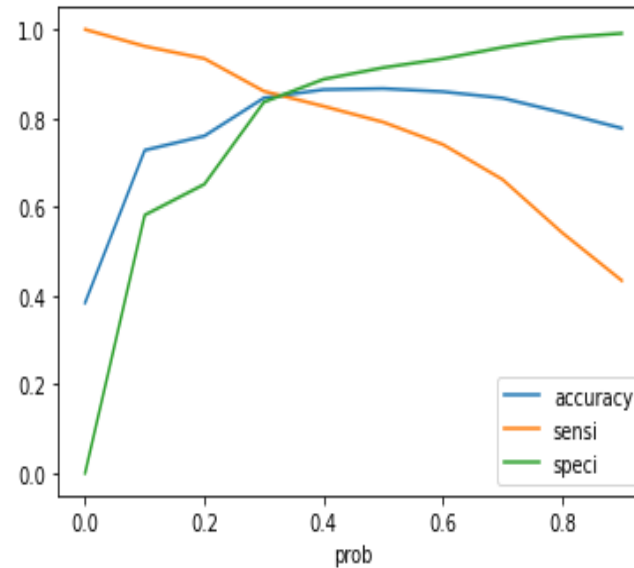
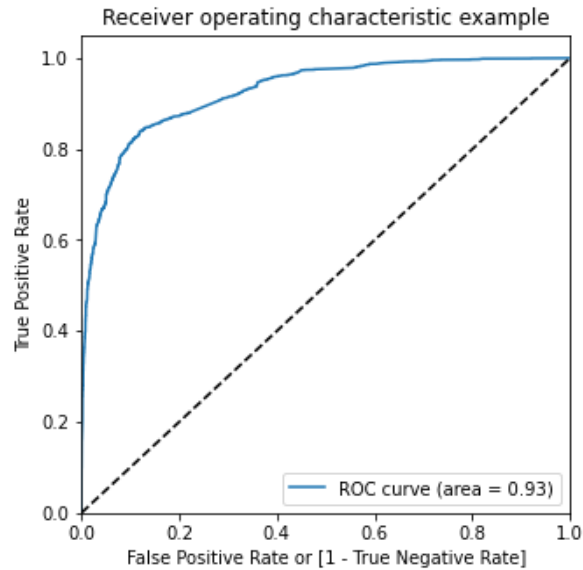


MODEL BUILDING

- Splitting the Data into Training and Testing Sets (70% train and 30% test)
- In total 6 logistic regression (GLM from statsmodels) models are build
- Building Model by dropping features with high VIF (>5), and then dropped features with $p\text{-value} > 0.05$.
- Predictions on test data set
- Overall all metrics will give accuracy greater than 84%



FINDING OPTIMAL CUT OFF POINT



- **Finding Optimal Cut off Point**
- Op Area under the curve is 0.93
- The cut of point according to the 2nd graph (accuracy, sensitivity and specificity) is 0.34
- From the 3rd graph (precision recall graph) the cutoff point is 0.42



CONCLUSION AND RECOMMENDATIONS

It was found that the variables that mattered the most in the potential buyers are:

- The total time spend on the Website.
- When the lead source was:
 - Google
 - Direct traffic
 - Reference
- When the last activity was:
 - SMS
 - Olark chat conversation
- Customers with tags will revert after reading the email and Tags Closed by Horizzon has higher chances of conversion
- 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.

