Lead Scoring Case Study

BY:

Ashutosh Sharma & Rancy Chadha

Problem Statement & Objective

Problem Statement:

An education company named X Education sells online courses to industry professionals. X Education gets a lot of leads, its lead conversion rate is very poor.

The typical lead conversion rate at X education is around 30%. The company's CEO wants it to be increased to 80%

Objective:

Build a model to help X Education find hot and promising leads by assigning lead scores that can be used by their sales team to target the leads

Build a Logistic Regression model to find predict lead conversion probability

Assign a lead score between 0 to 100 to each of the leads to identify hot leads.
Lead score indicates chances of conversion such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

Approach

Data Cleaning and Manipulation
- Drop columns that have single values and redundant columns
- Handle columns with null/missing values by dropping or imputing
- Outlier Detection

Perform EDA
- Univariate and Bivariate Analysis
- Heatmap analysis

Data Preparation
- Convert Binary value columns
- Dummy Variable Creation
- Feature Scaling

Model Building
- Apply RFE to get top 15 best performing features
- Create Logistic Regression model
- Use high P-values and apply VIF to eliminate low performing features selected using RFE

Model Validation
- Using Sensitivity, Specificity, Pression and Recall identify optimal probability cut-off
- Apply model on test data

Lead Score – Assign lead score based on predicted probability multiplied by 100

EDA Analysis

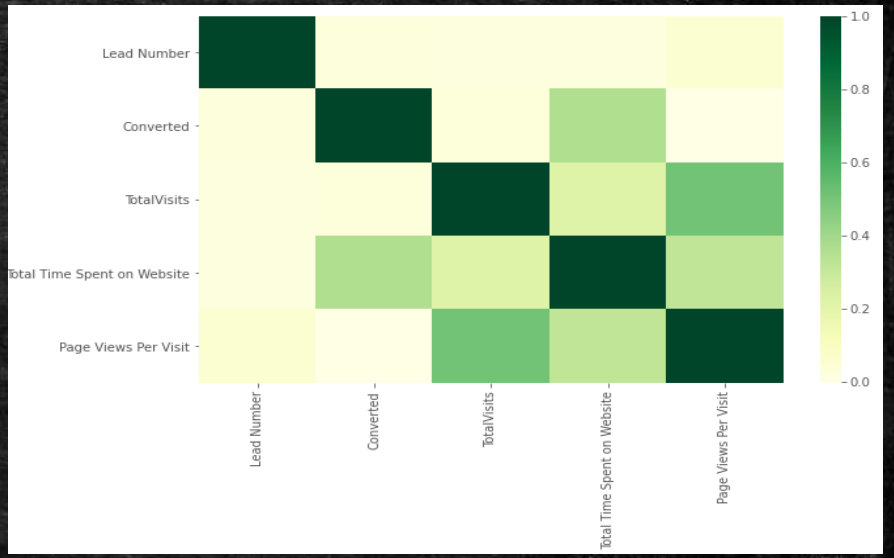
Google is the Lead Source that showed maximum successful Lead Conversion

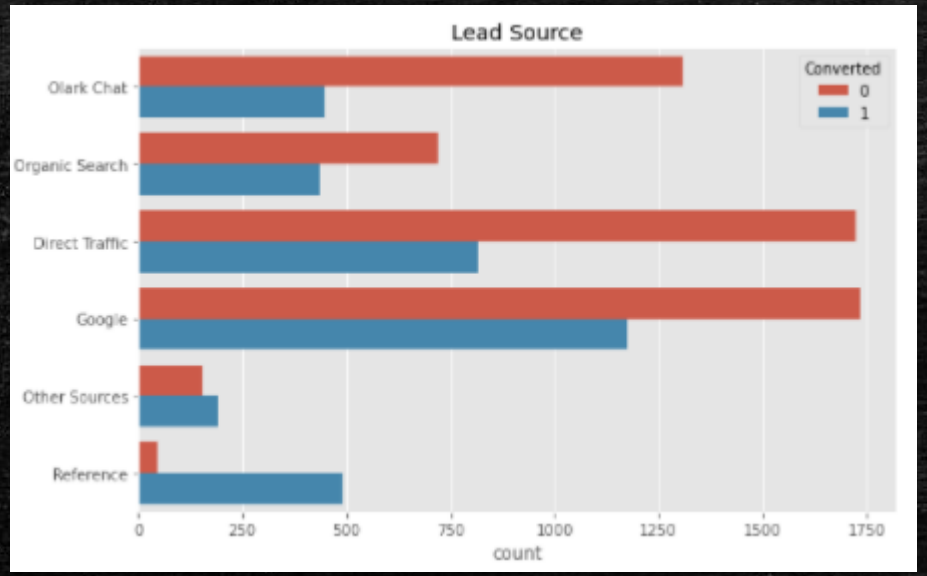
Under Lead Origin 'Landing Page Submission' had major lead conversion

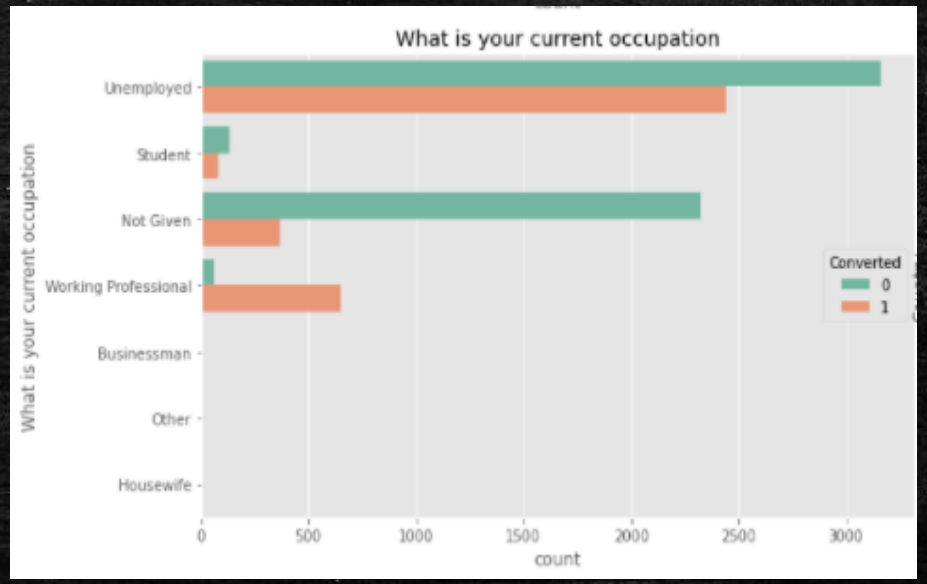
While Unemployed under What is your current occupation enquire the most, but their conversion rate is around 60-65% whereas working Professionals have highest rate of conversion

From heatmap we see Converted and Total time spent on website have good correlation

Total visits and Page Views Per Visit seems to have no correlation with Converted



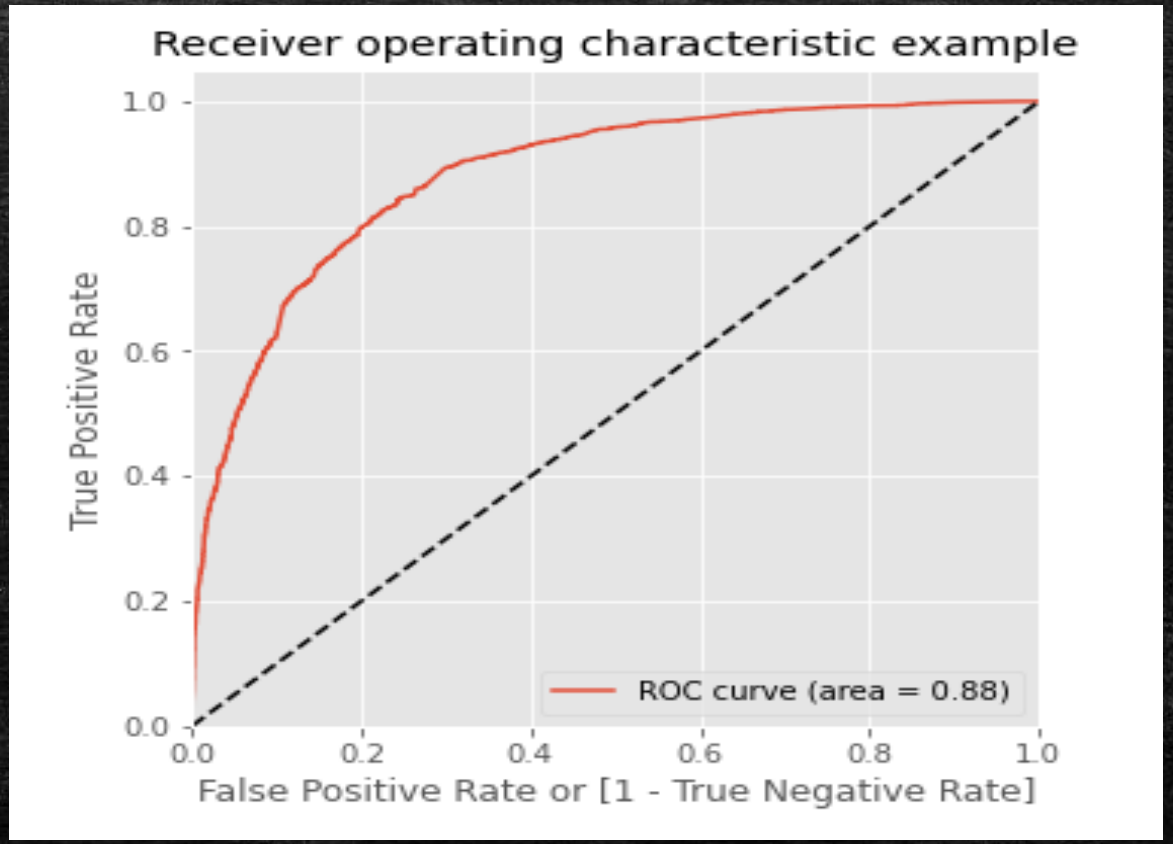




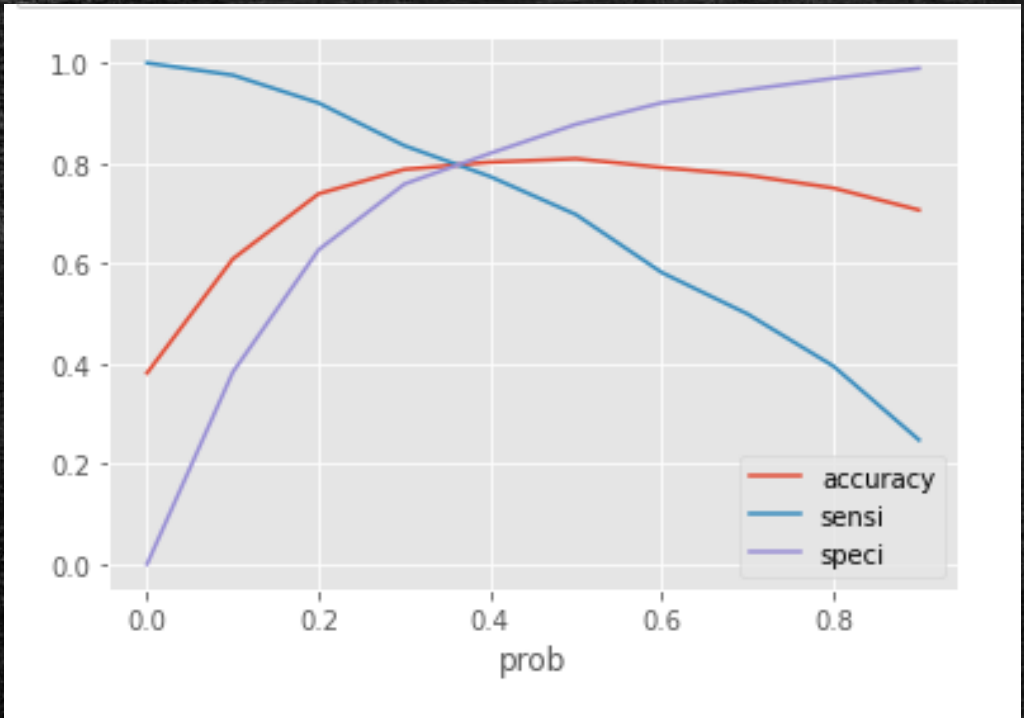
ROC Curve

The probability of the lead conversion appears to be quite high.

As the area under ROC Curve of the covers 88%



Optimal Probability Cut Off



The primary requirement is to understand at what probability would we consider a lead as a Hot Lead.

We can see from the visual plotting accuracy, sensitivity and specificity that 0.37 can be considered as the cut-off

Thus, any lead having probability of conversion of 0.37 or above can be safely considered as a hot lead to pursue. Therefore X Education company should go after leads with above 0.37 probability

Model Performance Statistics

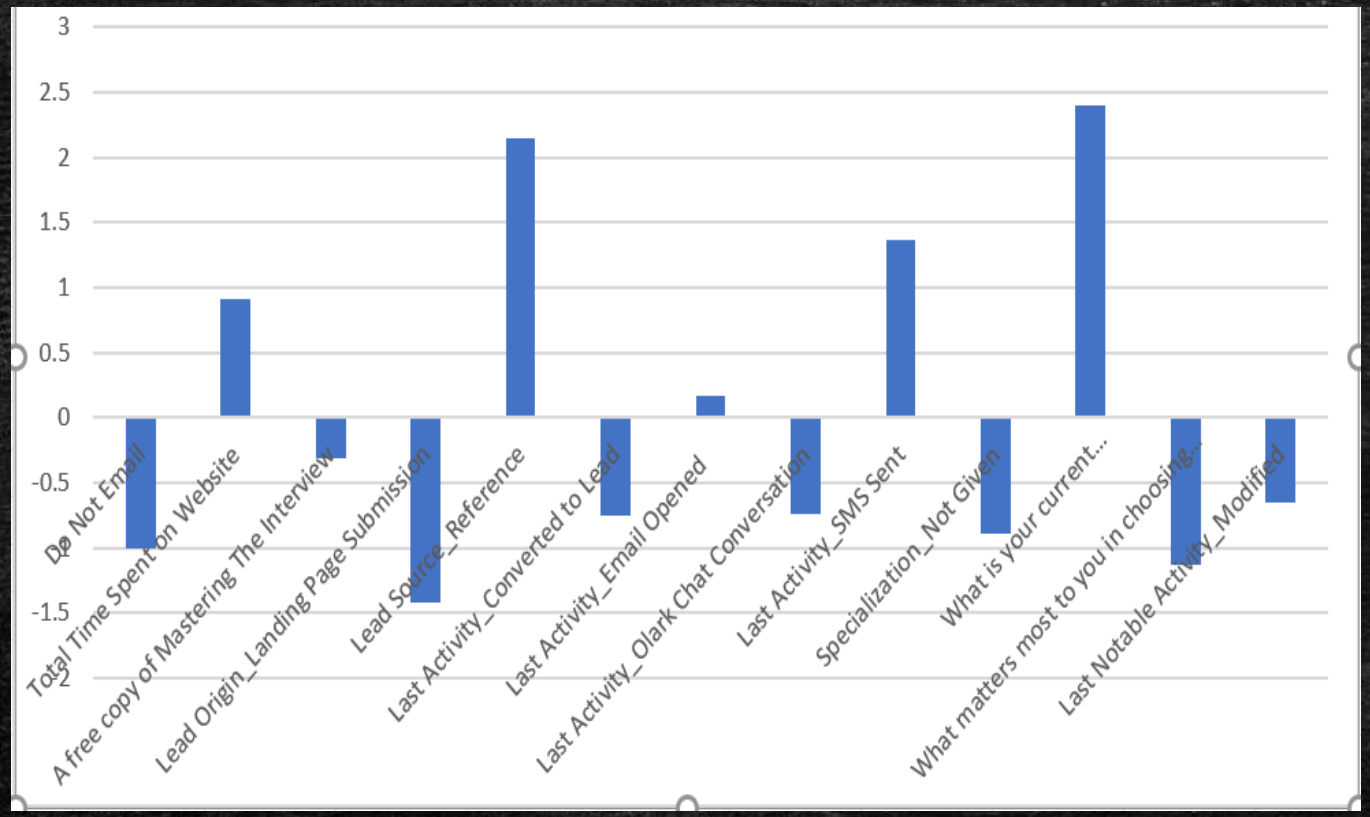
Confusion Matrix

Predicted/Actual	Not Converted	Converted
Not Converted	1285	192
Converted	196	899

We can see that model has sensitivity of 82.1% which means historic leads could have got converted with 82% chances

The model can correctly predict with 78.8% Accuracy and 69% precision

Top Impactful Features for Lead Conversion



Top 3 variables in the model which contribute most towards the probability of a lead getting converted are:

- **What is your current occupation_Working Professional**
- **Lead Source_Reference**
- **Last Activity_SMS Sent**

Top 3 variable which needs to be focused the most to increase the probability of lead conversion

- **Lead Origin_Landing Page Submission**
- **Specialization_Not Given**
- **What matters most to you in a course_Not Given**

Recommendations

- Any lead beyond 0.37 cutoff probability should be considered as a Hot Lead
- Sales teams should focus on customer that are provided via referrals as Lead Source_Reference is the lead source that provided potential customers
- Working professionals as per the model are another type of customers that should be focused for lead conversions
- Leads that are generated from Landing Page Submissions can be given lesser priority while looking for potential customers
- People who have opted for Do Not Email can also given minimal priority for lead conversion