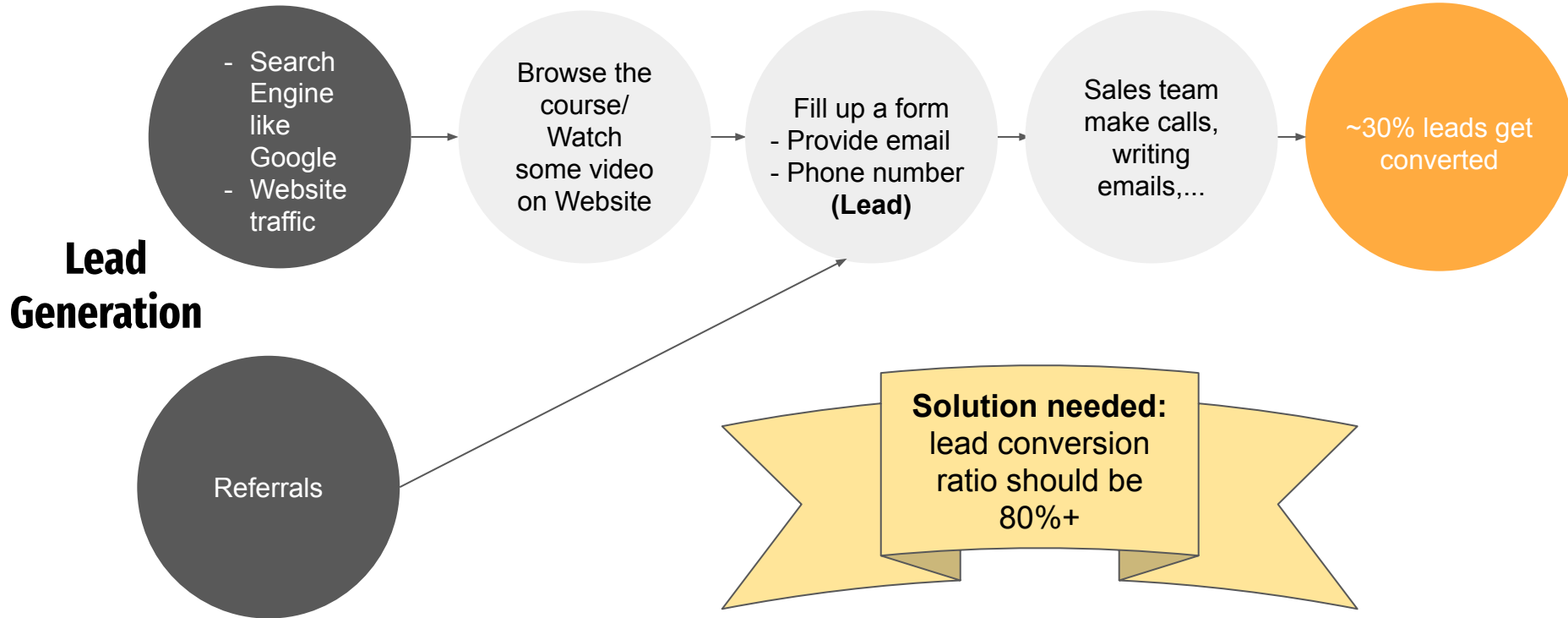# Lead Scoring Case Study

Thach Thao Ho | DSC65

Problem Statement

- X Education sells online courses for industry professionals.
- The company gets many leads, but the conversion rate is low.
- Only 30% become paying customers.
- To improve this, the company wants to identify the most promising leads, called 'Hot Leads.'
- By focusing on 'Hot Leads,' the sales team can improve the conversion rate by spending more time on likely customers.

Lead - Conversion Process

Lead Generation

- Search Engine like Google
- Website traffic

Browse the course/ Watch some video on Website

Fill up a form
- Provide email
- Phone number **(Lead)**

Sales team make calls, writing emails,...

~30% leads get converted

Referrals

Solution needed: lead conversion ratio should be 80%+

Solution

1. Narrowing down the selection of hot leads.

We group the leads into specific categories based on their likelihood to convert, allowing us to narrow down and focus more on a smaller group of hot leads.
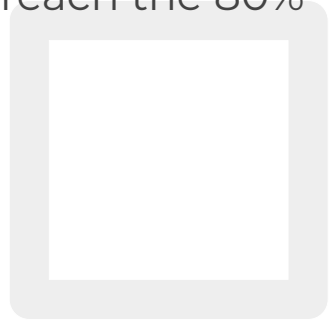
2. Focus on communicating with the potential leads

With a smaller set of leads to engage with, we can have a greater impact through more effective communication.

3. Convert

By focusing on hot leads, which are more likely to convert, we can improve the conversion rate and reach the 80% target.

* The need to target accurate hot leads is important.

Execution

Step 1: Data Cleaning and Manipulation

1.1. Handle Duplicates: Check and remove duplicate data entries.

1.2. Address Missing Values: Identify and handle NA and missing values.

1.3. Drop Unnecessary Columns: Remove columns with too many missing values that are not useful for analysis.

1.4. Impute Values: Fill in missing values where necessary.

1.5. Manage Outliers: Identify and handle outliers in the data.

Result - Data Cleaning and Manipulation

- The dataset contained 37 rows and 9240 columns.
- Features with a single unique value: "Magazine," "Receive More Updates About Our Courses," "Update My Supply Chain Content," "Get Updates on DM Content," and "I Agree to Pay by Cheque," were dropped.
- "ProspectID" and "Lead Number," which were not relevant to the analysis, were also dropped.
- Features with minimal variation: "Do Not Call," "What Matters Most to You in Choosing a Course," "Search," "Newspaper Article," "XEducation Forums," "Newspaper," and "Digital Advertisement," were dropped.
- Columns with over 35% missing values: 'How Did You Hear About X Education' and 'Lead Profile,' were dropped.

Step 2: Exploratory Data Analysis (EDA)

1. Univariate data analysis: value count, distribution of variables, etc.

2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.

3. Feature Scaling & Dummy variables and encoding of the data.

4. Classification technique: logistic regression is used for model making and prediction.

5. Validation of the model.

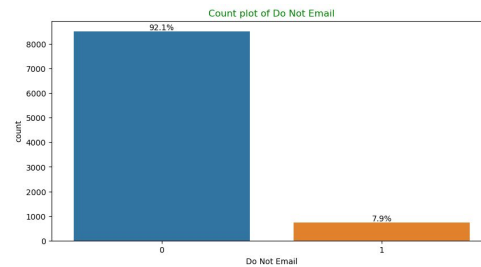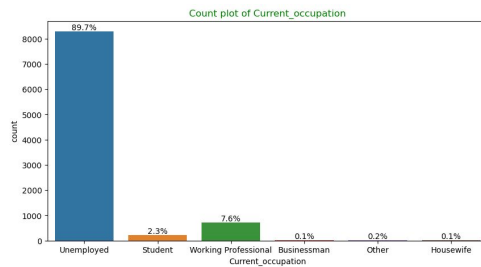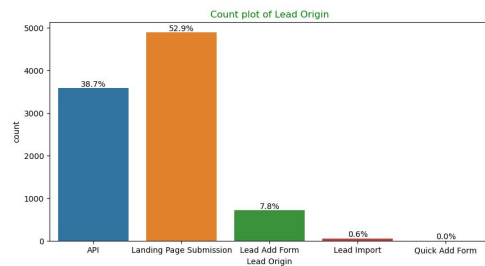6. Model presentation.

7. Conclusions and recommendations.

Leads Converted
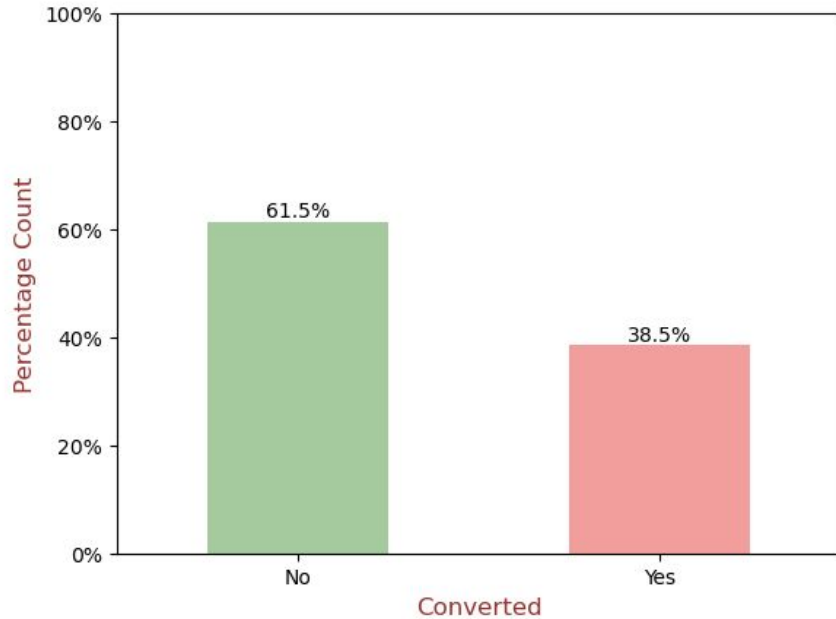
38.5% of people successfully converted to leads, whereas 61.5% did not.

Count plot of Lead Origin

Count plot of Current_occupation

Count plot of Do Not Email

- **Lead Origin**: 52.9% of customers came from "Landing Page Submission," while 38.7% came from the "API."
- **Current Occupation**: 90% of customers are unemployed.
- **Do Not Email**: 92% of individuals have declined to receive emails about the course.

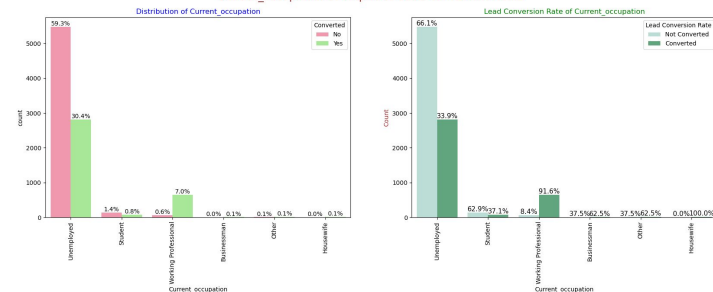Count plot of Lead Source

Count plot of Last Activity
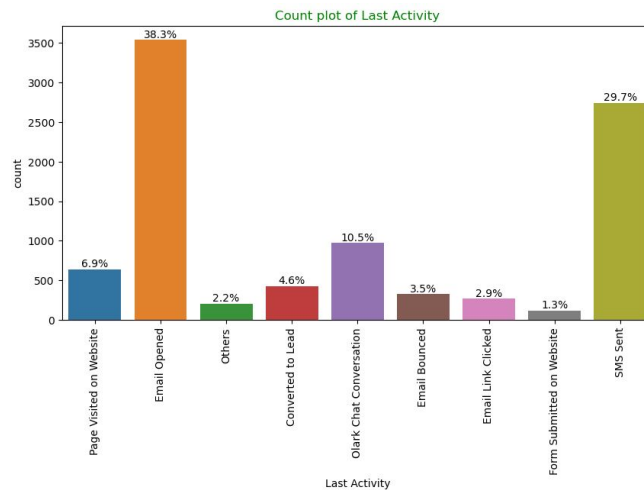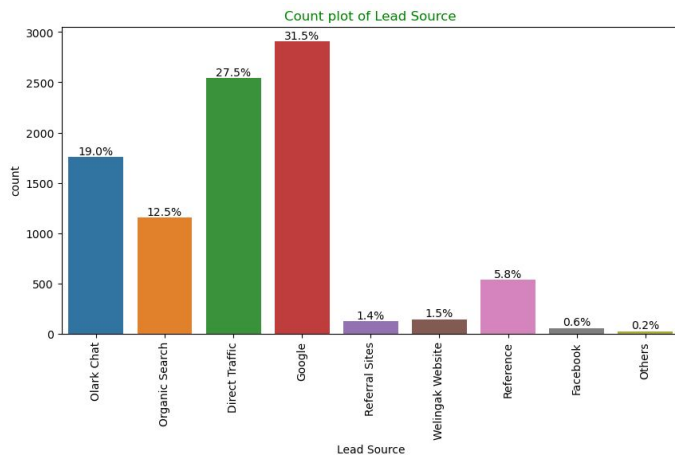
- **Lead Source**: 59% of leads originate from both Google and Direct Traffic.
- **Last Activity**: SMS Sent and Email Opened makeup 68% of customer activity.

- **Lead Origin**: Approximately 52% of leads came through "Landing Page Submission," with a conversion rate of 36%. The "API" identified about 39% of customers, with a 31% lead conversion rate.
- **Current Occupation**: Around 90% of customers are unemployed, with a 34% conversion rate. On the other hand, working professionals, who represent only 7.6% of the total customers, have a high 92% conversion rate.
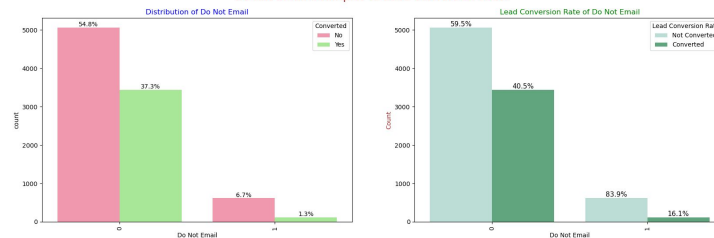- **Do Not Email**: 92% of individuals opted not to receive emails about the course.

Lead Origin Countplot vs Lead Conversion Rates
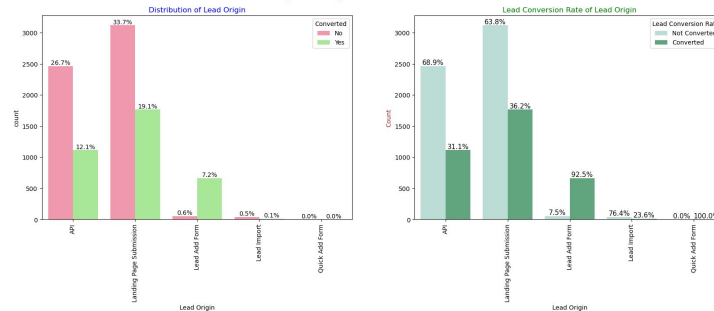
Current_occupation Countplot vs Lead Conversion Rates

Do Not Email Countplot vs Lead Conversion Rates

- **Lead Source**: Google has a 40% lead conversion rate out of 31% of customers. Direct Traffic contributes a 32% Lead Conversion Rate with 27% of customers, which is lower than Google. Organic Search provides a 37.8% Lead Conversion Rate, but only 12.5% of customers come from this source. Reference has an Lead Conversion Rate of 91%, but only around 6% of customers come through this lead source.
- **Last Activity**: "SMS Sent" has a high lead conversion rate of 63%, with 30% of contributions from last activities. The "Email Opened" activity contributed to 38% of the last activities performed by customers, with a 37% lead conversion rate.
- **Specialization**: Specializations in Marketing Management, HR Management, and Finance Management show good contributions.

Lead Source Countplot vs Lead Conversion Rates

Last Activity Countplot vs Lead Conversion Rates

Count plot of Specialization
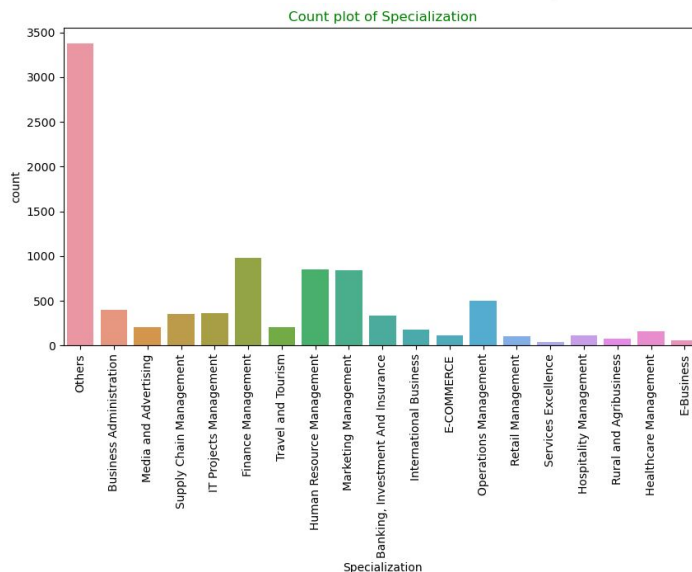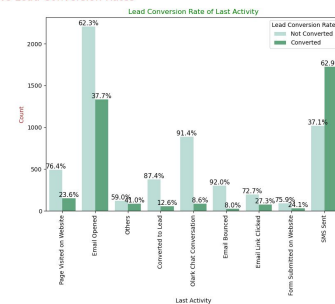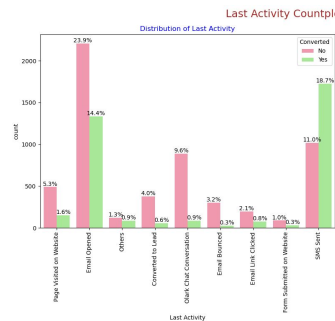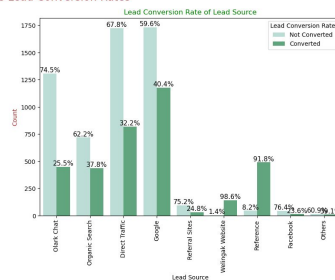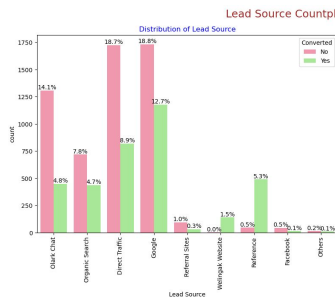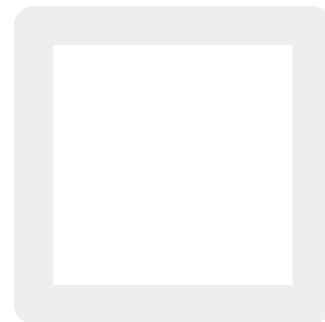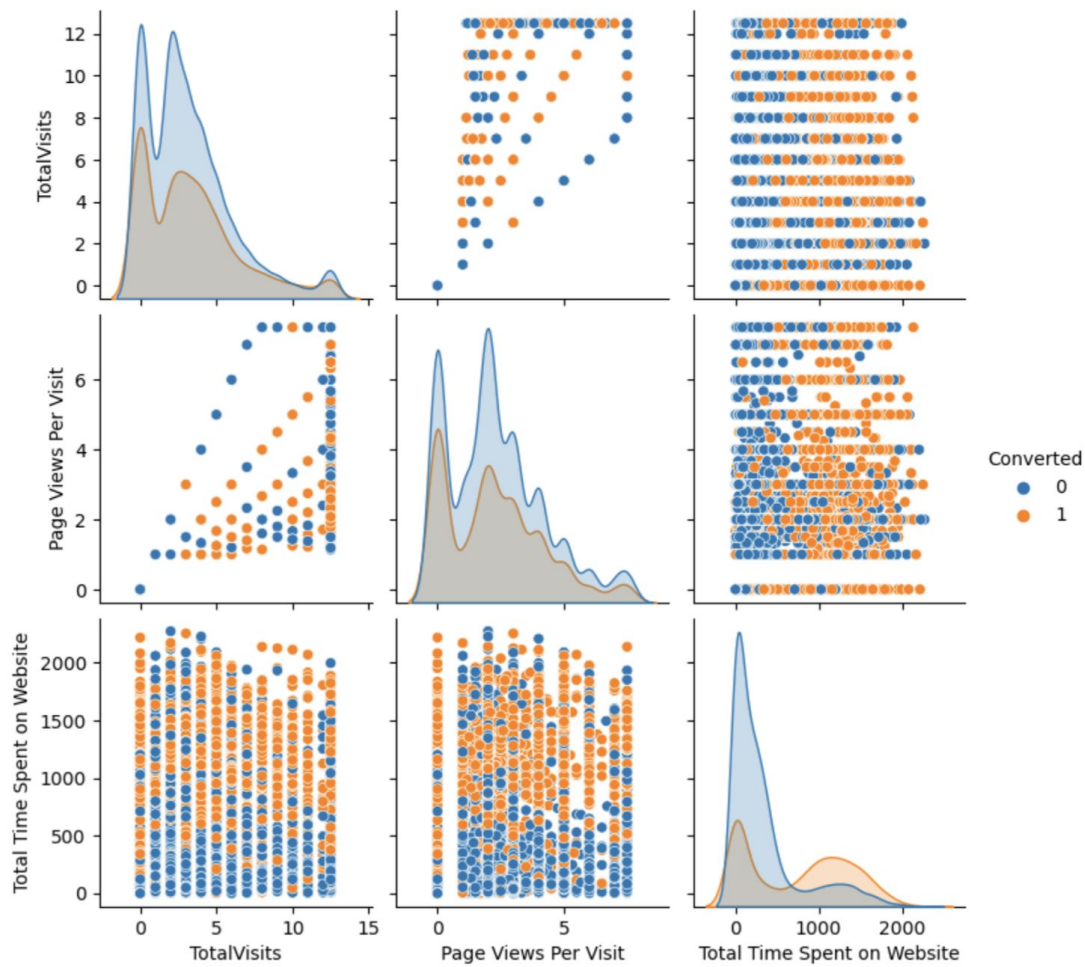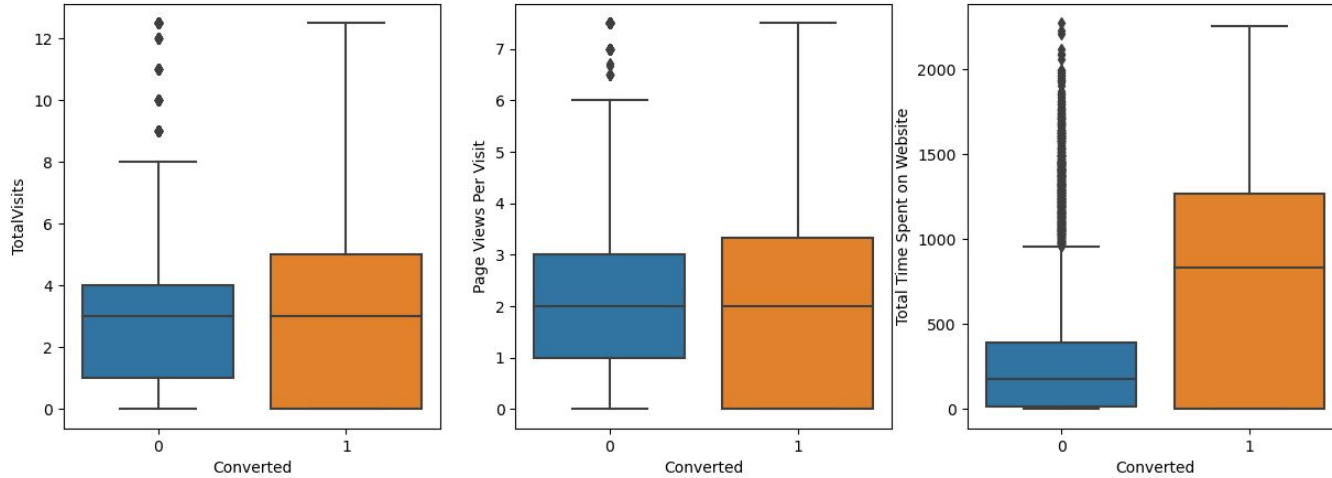
Past Leads who spends more time on Website are successfully converted than those who spends less as seen in the boxplot

Step 3: 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 vi value is greater than 5

Predictions on test data set Overall accuracy 81%

- Finding Optimal Cut off Point
- Optimal cut-off probability is that
- Probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.35.

Receiver operating characteristic example
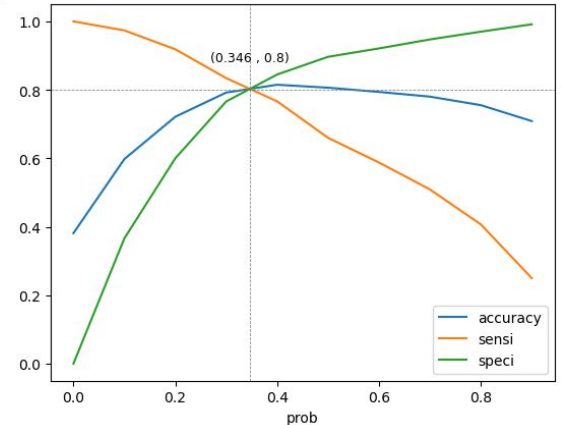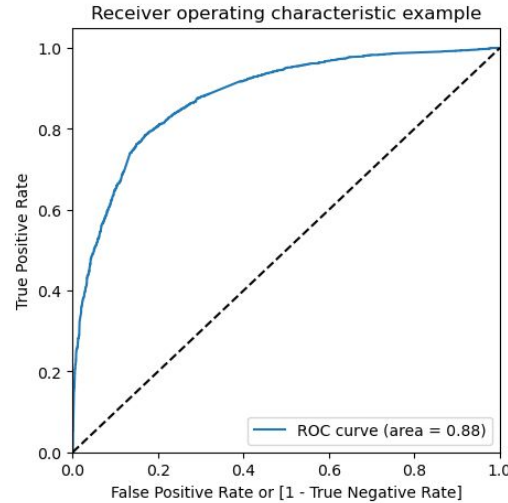
Test set Prediction

- Accuracy : 80.34%
- Sensitivity : 79.54%
- Specificity : 80.86%
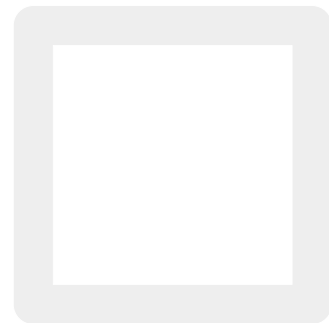
Conclusion

The key factors that drive potential buyers, ranked by importance, include: the total time they spend on the website, the number of visits, their lead source (Google, direct traffic, organic search, Welingak website), their last activity (SMS, Olark chat), lead origin (Lead add format), and whether they are working professionals.

To improve the overall conversion rate, X Education Company needs to concentrate on these critical factors:

- Increase total visits through advertising efforts to raise the conversion rate.
- Strengthen website user engagement to achieve better conversion results.
- Focus on sending more SMS notifications, which contribute to higher conversions.
- Improve the Olark Chat service, which is currently hindering conversions.

Thank You!