Leads Scoring
Case Study

SUBMITTED BY:

RINKLE DAS
SAIF ALI KHAN
RUSHITHA PALAKONDA

INDEX

Problem Statement

X Education gets a lot of leads from several websites and search engines like Google, once people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When people fill up a form providing their email address or phone number, they are classified to be a lead.

It's 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.

Objectives of Study

There are quite a few goals for this case study.

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

- There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

Data Cleaning

Handling Null Values

The columns with more than 35% null values are dropped.

Asymmetrique Profile Index (45.65%)
Asymmetrique Activity Index (45.65%)
Lead Profile (74.19%)
How did you hear about X Education(78.46%)

Imputed missing values

The columns with Null values replaced with 'Not Provided' instead of removing rows in the data

Specialization (36.58 %)
What matters most to you in choosing a course (29.32 %)
Country (26.63 %)

# Exploratory Data Analysis

Univariate Analysis

People with Free copy of the Mastering the Interview not required is higher than the people who are looking for Free copy of the Mastering the Interview

Also more number of people doesn't email with the course details are than wanting to share the course details through mail.
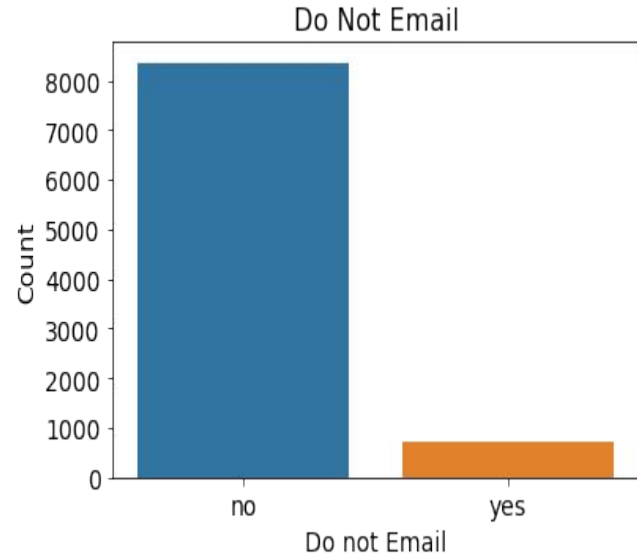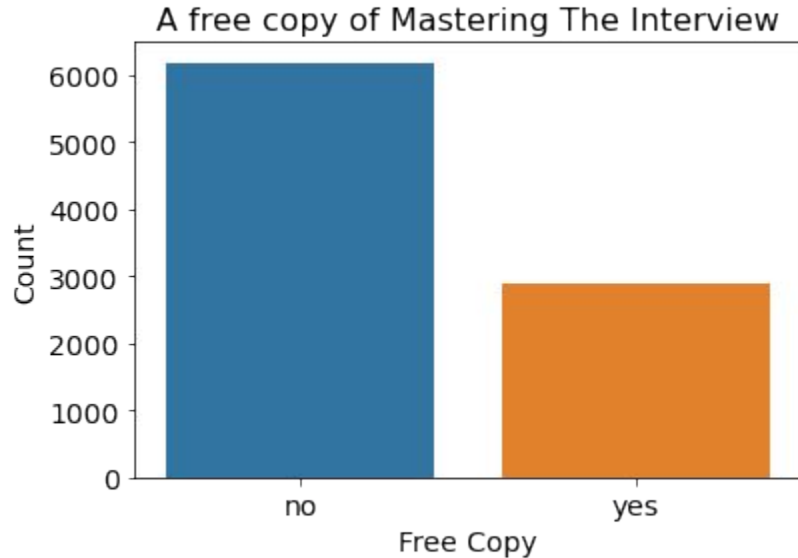
Modified and Email Opened is the highest last notable activity performed compared with other activities.

Landing Page submission is the highest count in lead origin compared with other Lead Origins.
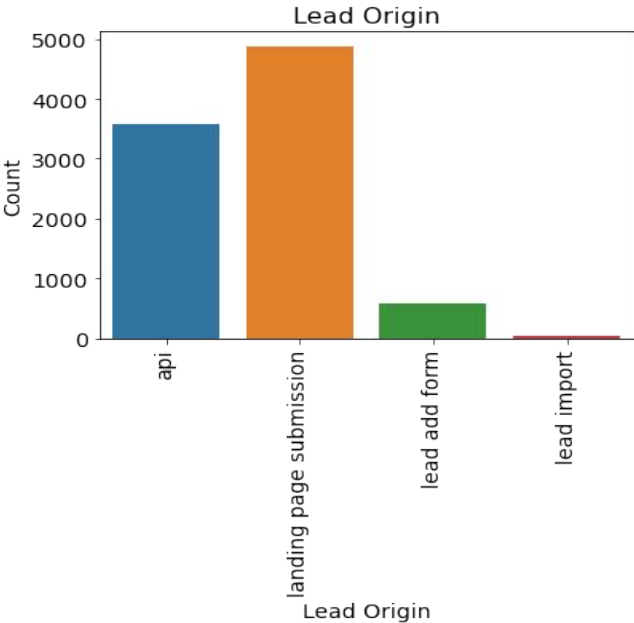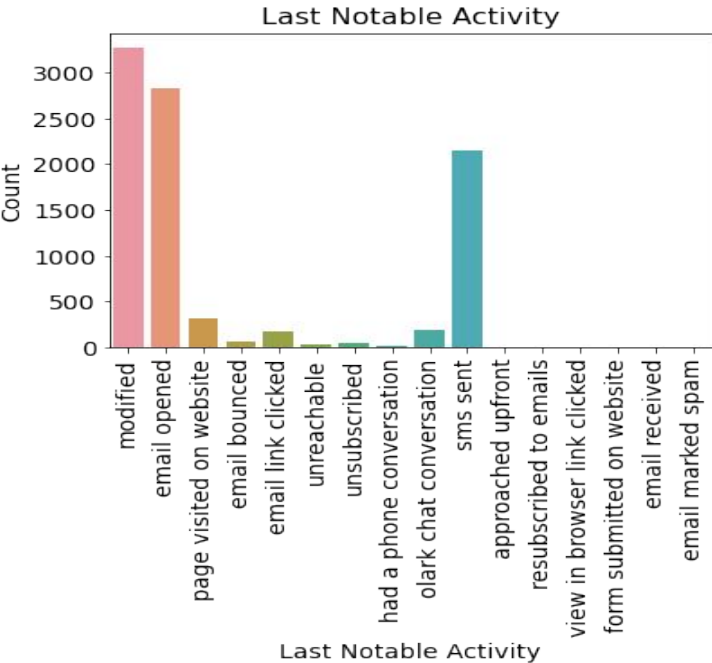
Bivariate Analysis

Upon performing a correlation matrix
test to understand the strength
between the variables, the EDA
reveals that several elements have
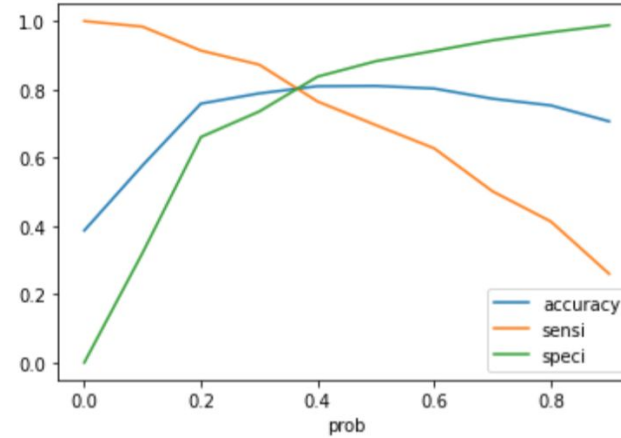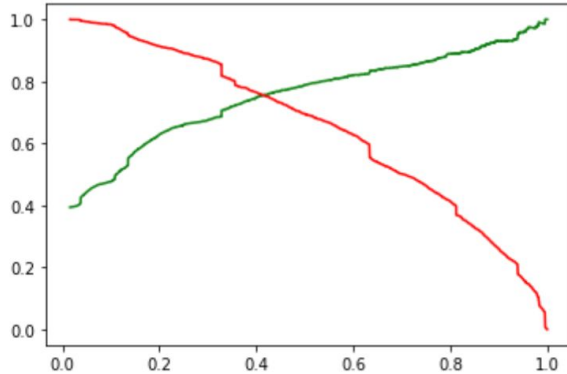minimal data and are likely less
relevant to the analysis.

Among the variables analysed we see
that Page views per visit ~ Total visits
and Total visits ~ Page views per visit
seem to have the strongest
correlation.
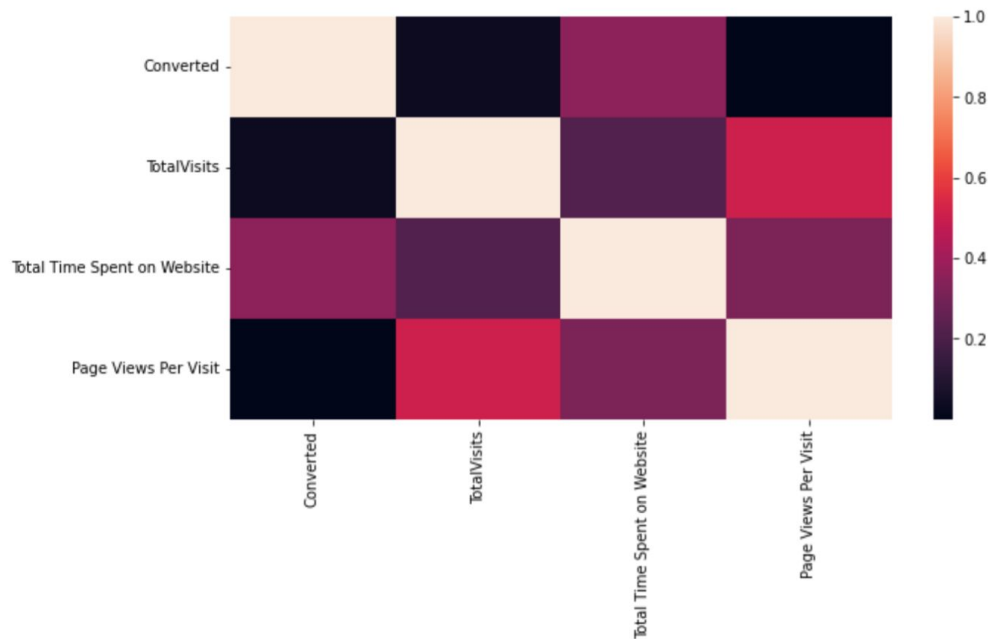
Model Details

Model Evaluation

At the cutoff of 0.35, the model demonstrates an accuracy, sensitivity, and specificity of approximately 80%.

Precision Re-call

At the current cutoff of 0.41, the model achieves a precision of approximately 75% and a recall of around 76%.

Conclusion

The analysis revealed that the key factors influencing potential buyers, ranked in descending order of importance, are:

1) Total time spent on the website
2) Total number of visits
3) Specific lead sources, including Google, direct traffic, organic search, and the Welingak website
4) Last activities such as SMS or Olark chat conversations
5) Lead origin being a Lead Ad form
6) Current occupation as a working professional.

By focusing on these critical aspects, X Education can effectively target and convert potential buyers, significantly increasing the likelihood of them enrolling in their courses.