LEAD SCORING CASE STUDY

SUBMITTED BY :

1. Darshan
2. Devika
3. Kamlesh
4. Damodar

TABLE OF CONTENTS

PROBLEM STATEMENT

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

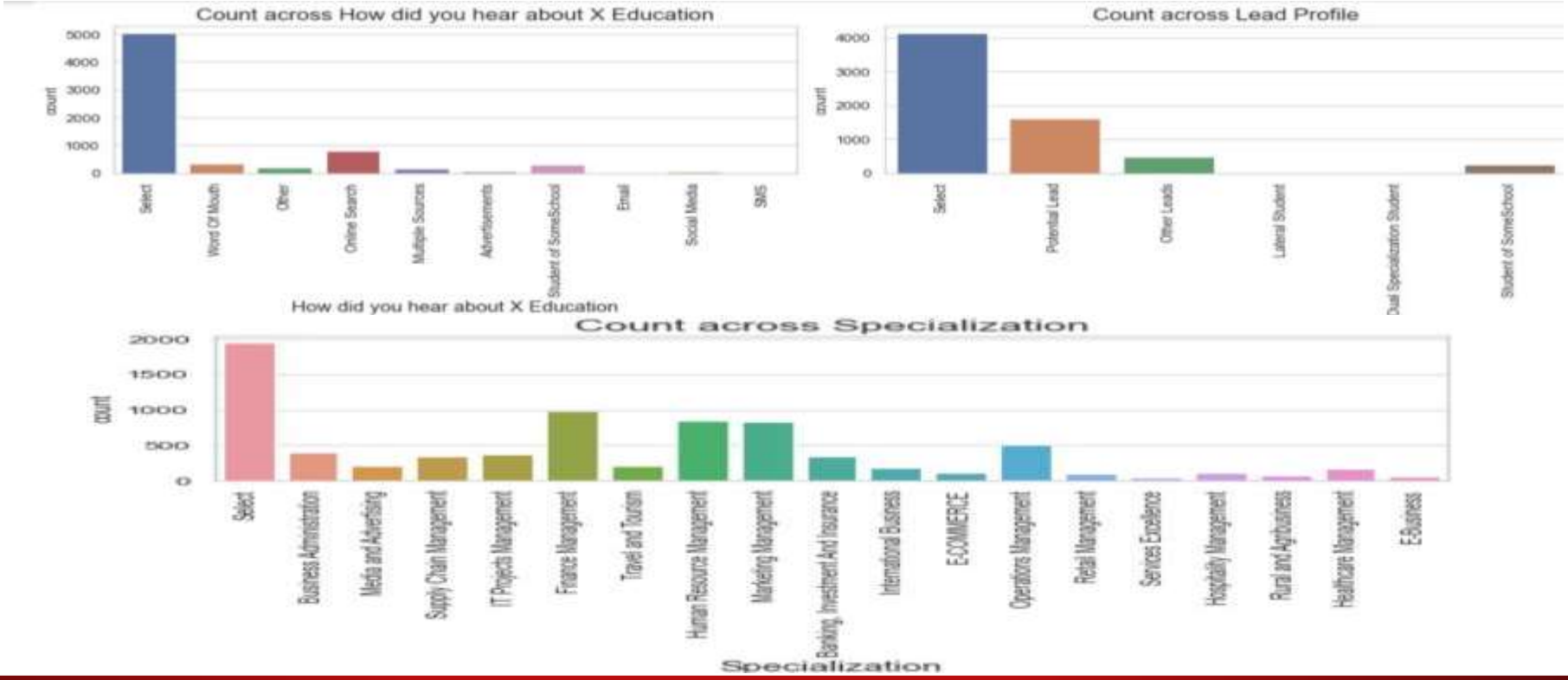
BUSINESS OBJECTIVE

Lead X has requested the development of a model that assigns a lead score ranging from 0 to 100 to each prospect. This scoring system is intended to pinpoint 'Hot leads', thereby enhancing the overall lead conversion rate. Additionally, the CEO aims to attain a remarkable lead conversion rate of 80%. The model should also be adaptable to future challenges, such as managing peak time activities, maximizing workforce utilization, and strategizing post-target achievement actions.

The process begins with importing the dataset and conducting an initial examination of the data frame. This is followed by preparing the data for analysis. An exploratory data analysis (EDA) is then performed to gain insights. Next, dummy variables are created for categorical data. The dataset is then divided into training and testing sets. Feature scaling is applied to normalize the data. The correlations between variables are analyzed. The model building phase includes utilizing Recursive Feature Elimination (RFE), assessing R-squared values, and checking VIF and p-values. The model is then evaluated for its performance. Finally, predictions are made using the test set.

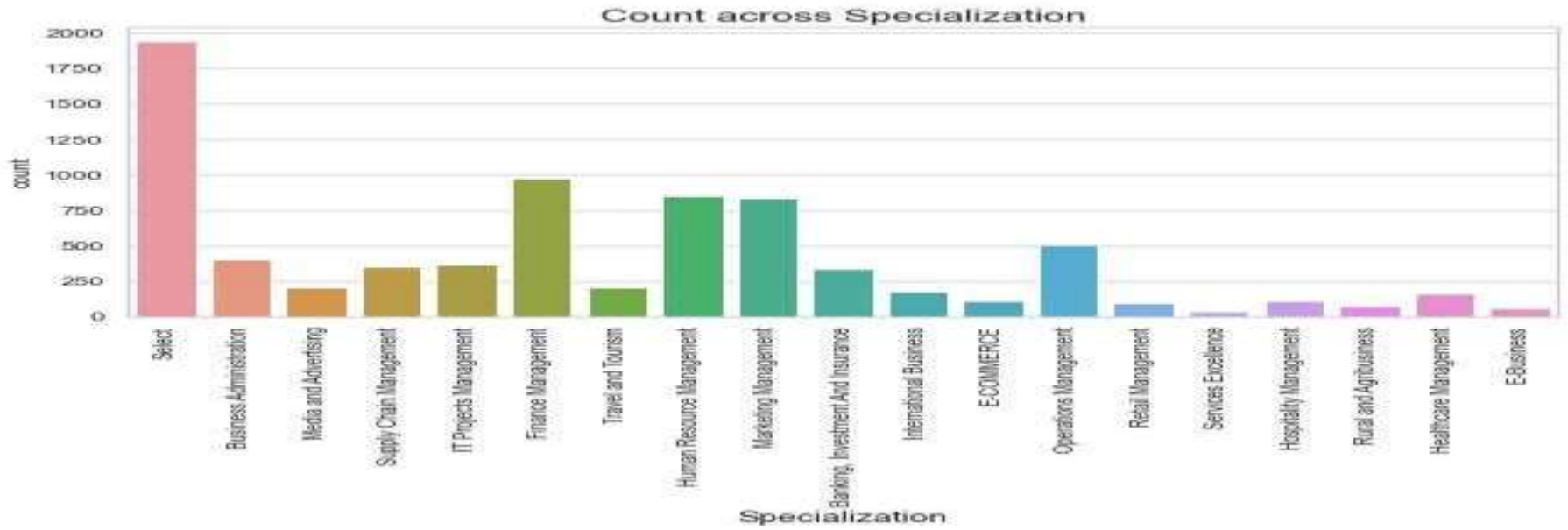
EDA – DATA CLEANING

- Several columns contain an option labeled 'Select', which is being addressed.



Specialization

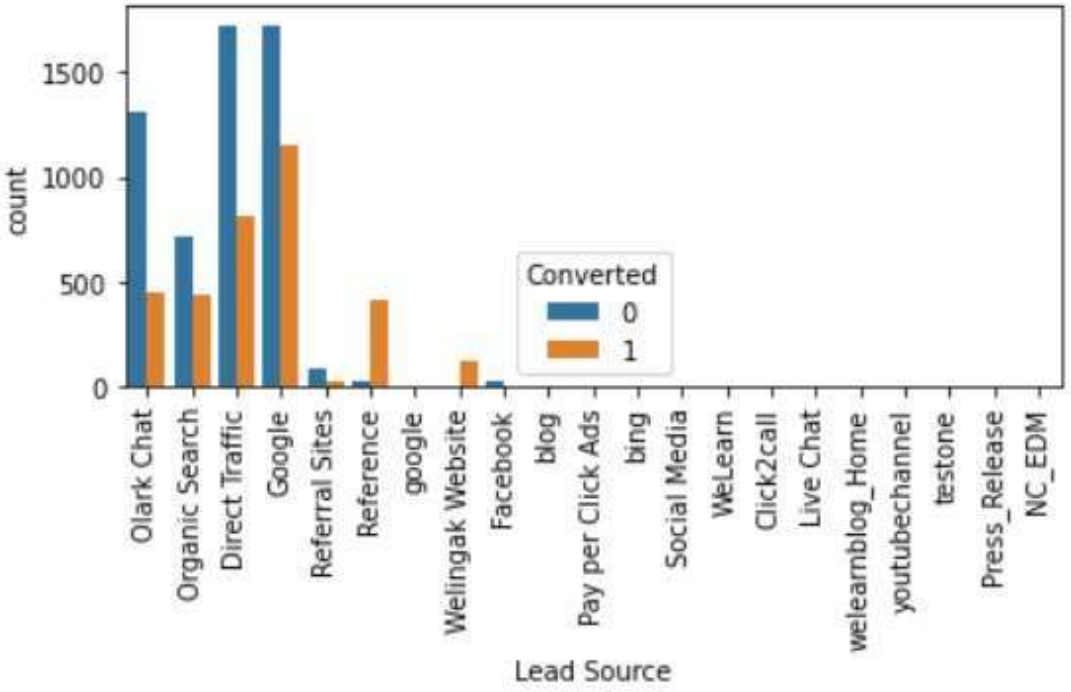
Leads specializing in the domains of Human Resources, Finance, and Marketing Management have a greater chance of successful conversion.

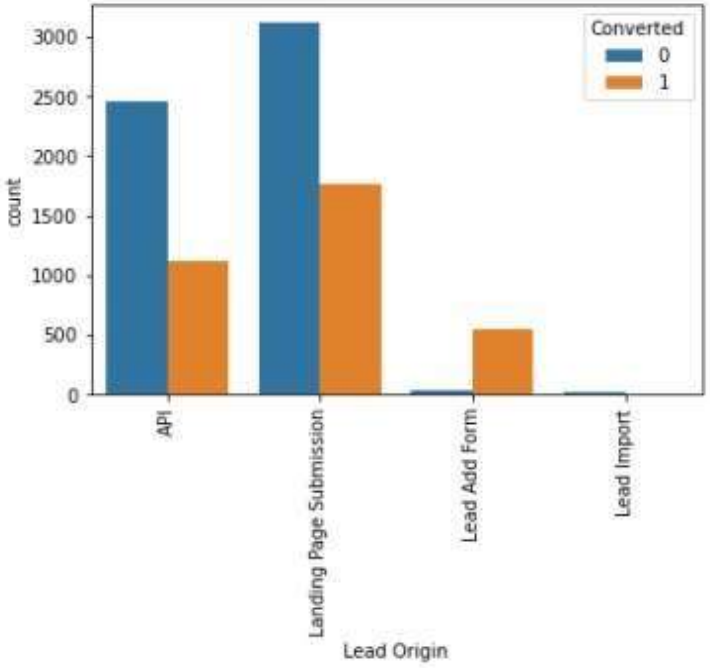


SOURCE & ORIGIN

In the analysis of lead sources, it's observed that leads acquired through Google and direct traffic have a high probability of conversion. This indicates that these channels are particularly effective in attracting potential customers who are more likely to complete the conversion process.

On the other hand, regarding the lead origin, it's noted that the majority of leads are coming from Landing Page Submissions. This means that the landing pages are a significant entry point for potential customers, drawing in a large number of leads.



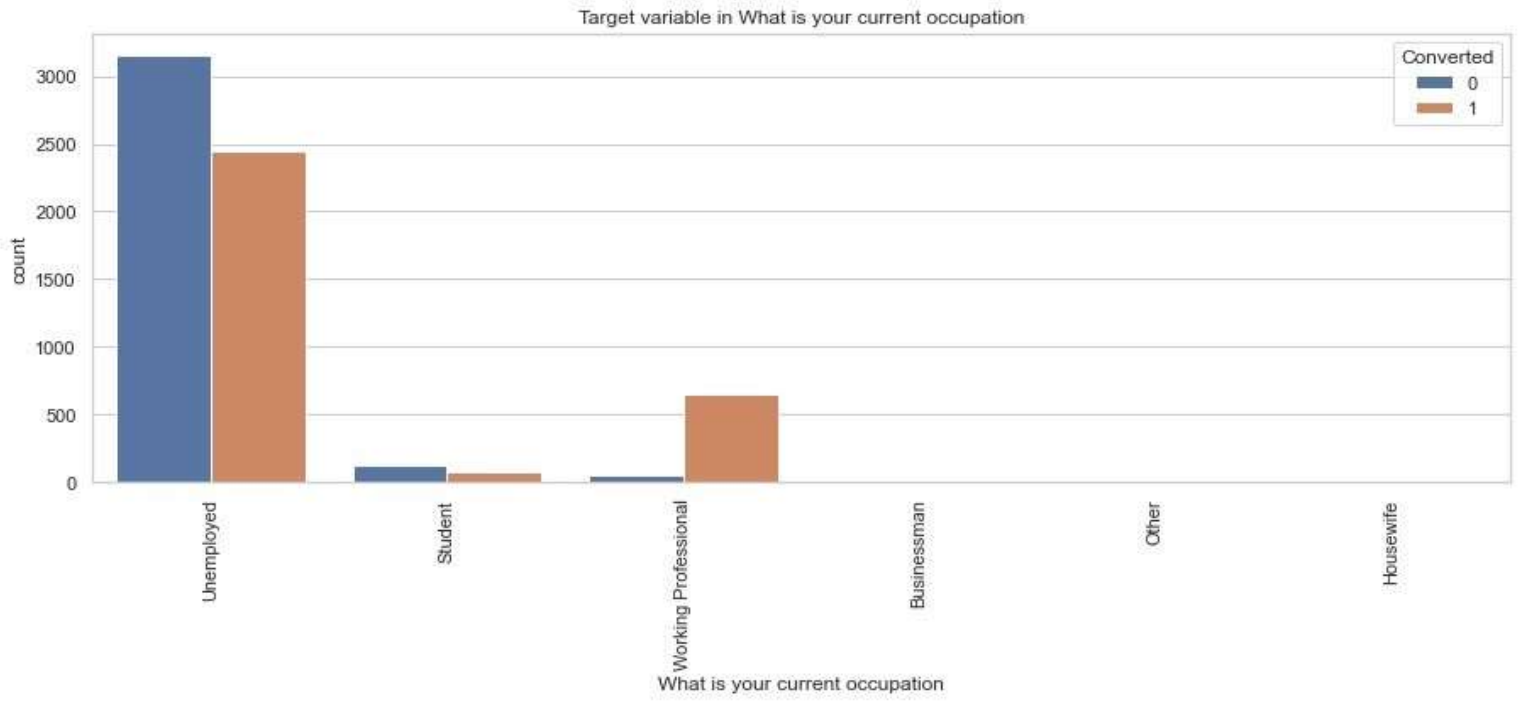


LAST LEAD ACTIVITY

LEADS THAT ENGAGE WITH EMAILS ARE MORE LIKELY TO CONVERT INTO CUSTOMERS. SIMILARLY, SENDING SMS MESSAGES IS ALSO ADVANTAGEOUS FOR CONVERSION. THIS IMPLIES THAT LEADS WHO ACTIVELY INTERACT WITH COMMUNICATION EFFORTS, SUCH AS OPENING EMAILS OR RESPONDING TO SMS MESSAGES, SHOW A HIGHER LEVEL OF INTEREST AND ENGAGEMENT. THESE ACTIONS INDICATE A STRONGER POTENTIAL FOR CONVERSION, MAKING SUCH LEADS PRIME CANDIDATES FOR FOCUSED MARKETING AND FOLLOW-UP STRATEGIES.

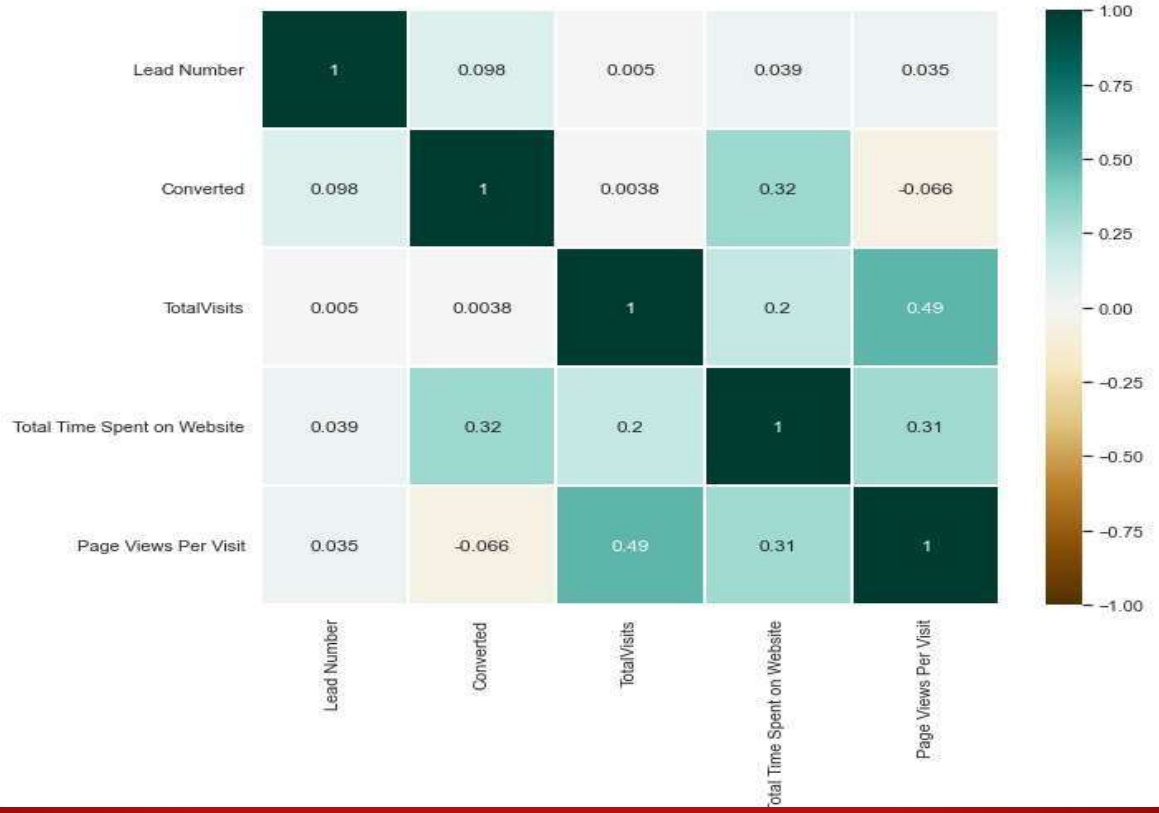
Occupation

Leads which are Unemployed are more interested to join the course than others



Correlation

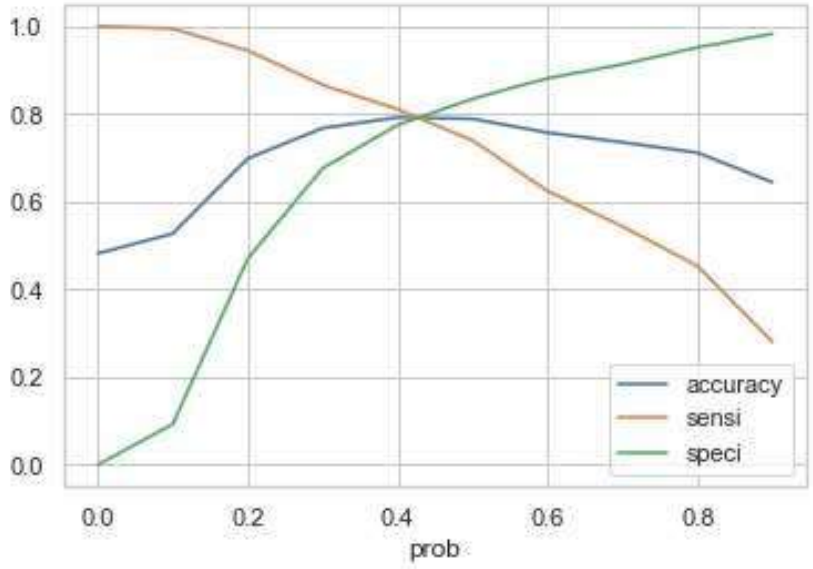
There is no correlation

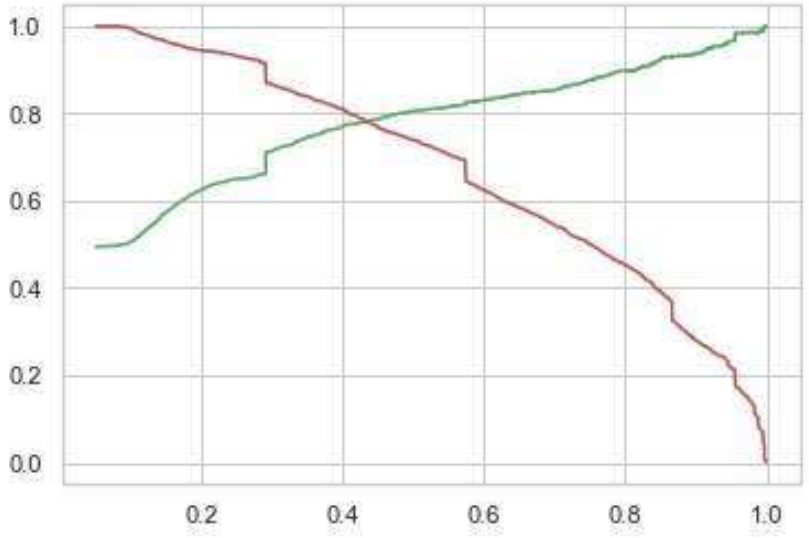


MODEL EVALUATION

0.42 is the tradeoff between Precision and Recall -

Thus we can safely choose to consider any Prospect Lead with Conversion **Probability higher than 42 % to be a hot Lead**





FINDINGS

For the training data set, the model achieved an accuracy rate of 80%, with a sensitivity (true positive rate) of 77% and a specificity (true negative rate) of 80%. These same performance metrics were mirrored in the test data set, with identical values for accuracy, sensitivity, and specificity.

Key features influencing these results include:

Leads originating from Olark Chat are a significant predictor.
This category, likely encompassing various unspecified specializations, plays a role in the model.
Leads generated through this form are crucial in the model's predictions.
Lead Source from the Welingak Website:This particular source of leads is notable for the model.
Total Time Spent on the Website:The duration a lead spends on the website is a critical factor.
Lead Origin from 'Landing Page Submission': This origin point for leads is another important predictor.

Two ther factors are influential:

The model places importance on whether a lead is a working professional.
The lead's preference regarding email communication is also a significant predictor in the model.

These factors combined contribute to the model's ability to accurately predict lead conversion, as reflected in the balanced accuracy, sensitivity, and specificity rates across both the training and testing datasets.

FINAL SAY

The data shows that leads generated through API and Landing Page Submission have a conversion rate of 30-35%, aligning with the average rate. In contrast, leads originating from the Lead Add Form and Lead Import have notably lower conversion rates. This suggests that focusing more on leads from API and Landing Page Submission could be more beneficial for conversion efforts.

The majority of leads come from Google or direct traffic sources. However, the highest conversion ratios are seen with leads referred from other sources or those coming from the Welingak website. This indicates that while Google and direct traffic generate the most leads, other sources like referrals and specific websites are more effective in terms of conversion.

Additionally, there's a clear trend showing that leads who spend more time on the website are more likely to convert, emphasizing the importance of website engagement in the conversion process.

Regarding lead activities, the most common last activity recorded is 'Email Opened', but the highest conversion rates are associated with leads who have been sent SMS messages. This highlights the effectiveness of SMS communication in lead conversion.

In terms of occupation, while most leads are unemployed, the highest conversion rates are seen with working professionals. This suggests that targeting employed individuals might yield better conversion results.

- **THANK YOU**