



Lead Scoring Case Study

APPROACH, FINDINGS AND RECOMMENDATIONS

Problem Statement

Identify leads which are more likely to convert using past conversion data of previous leads which were either converted or not converted.

- ▶ Build a Logistic Regression model which can predict the probability for a lead to be converted.
- ▶ Model should be interpretable to allow for identification of important variables which can be used as indicators for conversion.
- ▶ Solution and recommendations should be easily adjustable to accommodate the company's future requirement changes.

Problem Solving Approach

- ▶ Data Cleaning:
 - ▶ Handling missing values
 - ▶ Removing skewed variables
 - ▶ Identifying and handling outliers
- ▶ Exploratory Data Analysis(EDA):
 - ▶ Check target variable distribution
 - ▶ Analyse correlation between numerical fields
 - ▶ Visualize relationships between categorical variables and target variable
- ▶ Data Preprocessing:
 - ▶ Encoding categorical fields
 - ▶ Test-train split

Problem Solving Approach(contd.)

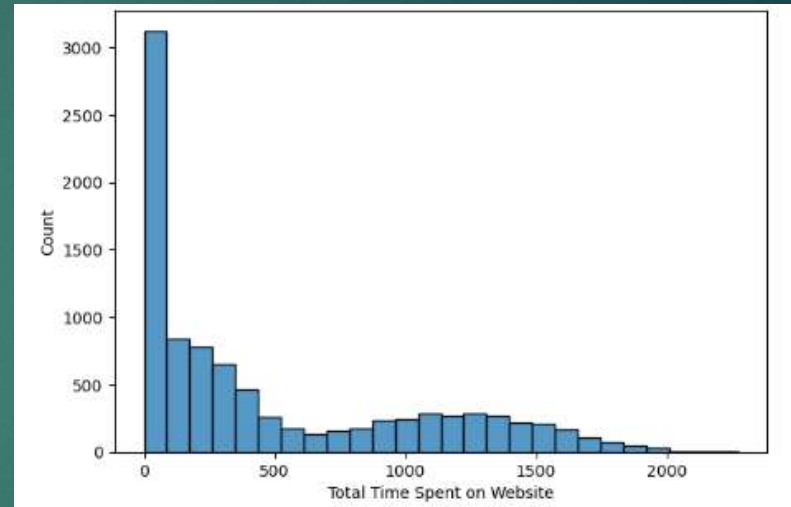
- ▶ Model Building:
 - ▶ Recursive feature elimination(RFE)
 - ▶ Iterative manual feature elimination
- ▶ Model Testing:
 - ▶ Selecting decision threshold
 - ▶ Evaluating model performance
- ▶ Interpretation and Results:
 - ▶ Analysing intercepts and coefficients
 - ▶ Interpreting the results

Top Variables from the Model

- ▶ **Total Time Spent on Website:** The total time spent by the customer on the website.
- ▶ **Lead Origin_Lead Add Form:** The origin identifier with which the customer was identified to be a lead. This is an encoded variable where the category was "Lead Add Form".
- ▶ **Lead Quality_Worst:** Indicates the quality of lead based on the data and intuition the employee who has been assigned to the lead. This is an encoded variable where the category was "Worst".
- ▶ **Last Notable Activity_SMS Sent:** The last notable activity performed by the student. This is an encoded variable where the category was "SMS Sent".
- ▶ **What is your current occupation_No Information:** Indicates whether the customer is a student, unemployed or employed. This is an encoded variable where the field was empty.

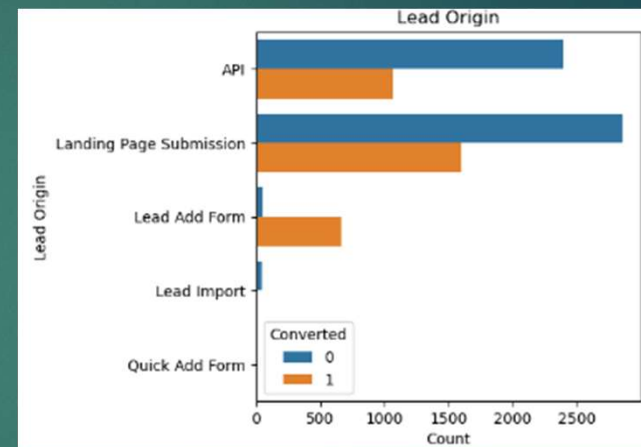
Top Variable: “Total Time Spent on Website”

- ▶ The top variable of the model
- ▶ The coefficient is positive
- ▶ This shows that the more time the lead spends on the website, they are more likely to convert
- ▶ If a lead has spent more time on the website, we need to focus on converting them



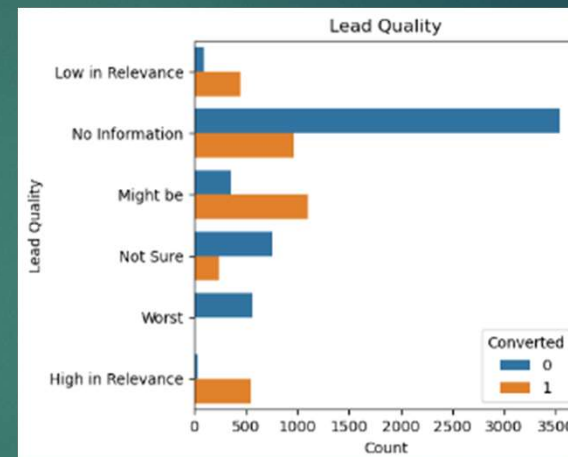
Top Variable: "Lead Origin_Lead Add Form"

- ▶ One of the top variables of the model
- ▶ The coefficient is positive
- ▶ This means that if the lead originated from "Lead Add Form", they are likely to convert
- ▶ If a lead originated from "Lead Add Form", we need to focus on converting them



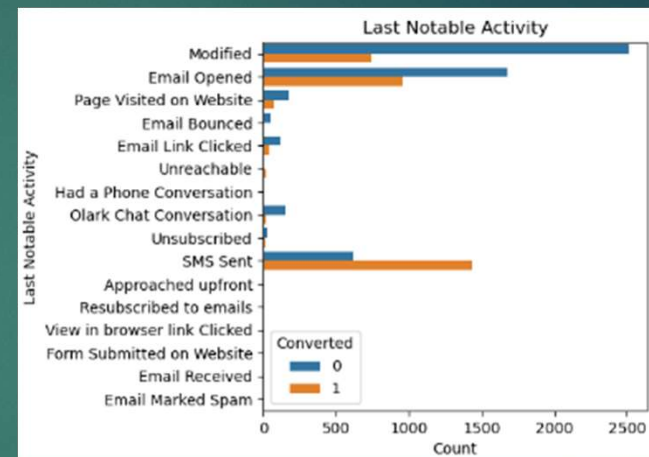
Top Variable: “Lead Quality_Worst”

- ▶ One of the top variables of the model
- ▶ The coefficient is negative
- ▶ This means that if the lead quality has been labelled as "Worst", they are less likely to convert
- ▶ If the lead has been labelled as "Worst", we should not focus on that lead



Top Variable: “Last Notable Activity_SMS Sent”

- ▶ One of the top variables of the model
- ▶ The coefficient is positive
- ▶ If the last activity of the lead was sending the SMS, they are likely to convert
- ▶ If the lead has sent the SMS, we need to focus on converting the lead



Top Variable: “What is your current occupation_No Information”

- ▶ One of the top variables of the model
- ▶ The coefficient is negative
- ▶ This is quite unique, as this shows that if the lead has not filled the ‘What is your current occupation’ field, they are less likely to convert
- ▶ If the lead doesn't take the time to fill in the information, they are less likely to convert
- ▶ It is assumed that the lack of information is a result of the lead not taking the time to fill in the information



Recommendations

- ▶ Prioritize leads who have spent the most time on the website/platform
- ▶ If the leads originate from “Lead Add Form”, we have to prioritize those leads as well.
- ▶ If the lead’s quality has been labelled as “Worst”, do not spend resources on trying to convert that lead.
- ▶ If “SMS Sent” was the last notable activity of a lead, focus efforts on trying to convert that lead
- ▶ If the lead has not filled the “What is your current occupation” field, do not prioritize the conversion of that lead