

Lead Scoring Case Study using logistic regression

SUBMITTED BY :

VINITA SHARMA

VINAYAK DHAMNEKAR

VINOD IYER

S ZAIBA

Content

- ▶ **Problem statement**
- ▶ **Analysis approach**
- ▶ **EDA**
- ▶ **Correlations**
- ▶ **Model Evaluation**
- ▶ **Business Results and Insights**
- ▶ **Summary**

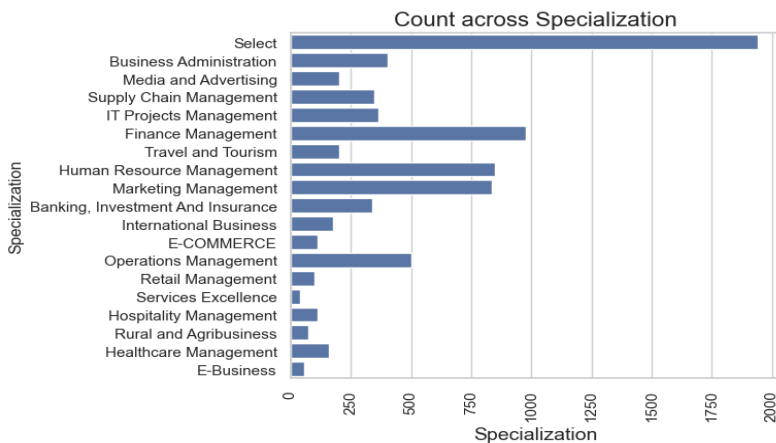
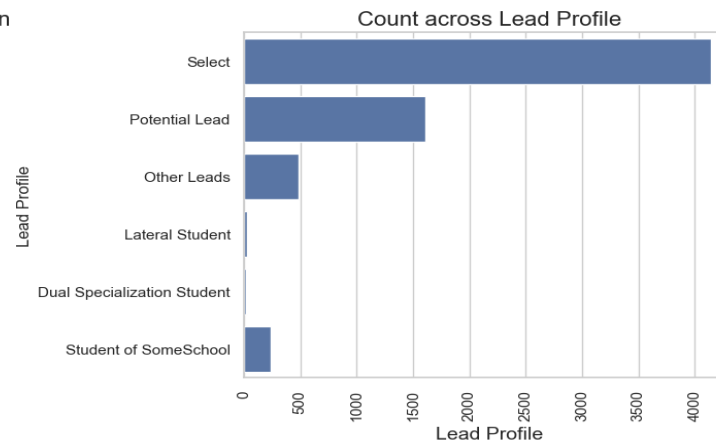
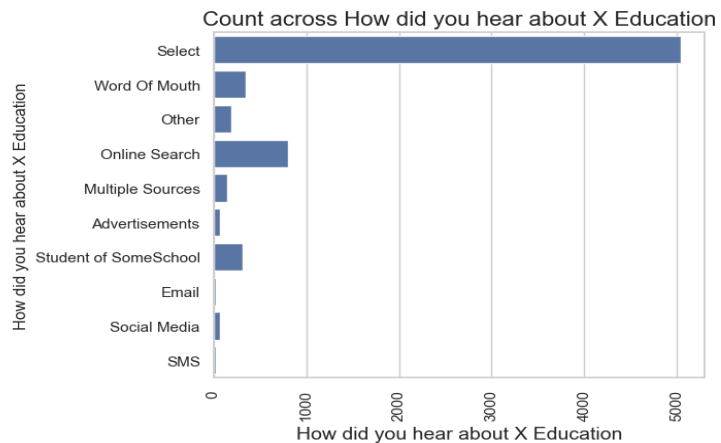
Problem Statement and Analysis Approach

- ▶ **Problem Statement:** The goal of the analysis is to identify the key factors contributing to lead conversion and to develop targeted strategies to optimize lead conversion efforts during different business phases.
- ▶ **Analysis Approach:**
 - ▶ **Variable Identification:** We identified and analyzed the top variables that significantly impact the probability of a lead getting converted.
 - ▶ **Categorical Variables:** We focused on key categorical/dummy variables that play a critical role in lead conversion.
 - ▶ **Strategic Recommendations:** Based on the analysis, we developed strategies for both aggressive lead conversion during high-resource periods (e.g., intern hiring phase) and minimizing unnecessary phone calls during slow periods.

Problem Approach

- ▶ **Importing the data and inspecting the data frame**
- ▶ **Data preparation**
- ▶ **EDA**
- ▶ **Dummy variable creation**
- ▶ **Test-Train split**
- ▶ **Feature scaling**
- ▶ **Correlations**
- ▶ **Model Building (RFE Rsquared VIF and p- values)**
- ▶ **Model Evaluation**
- ▶ **Making predictions on test set**

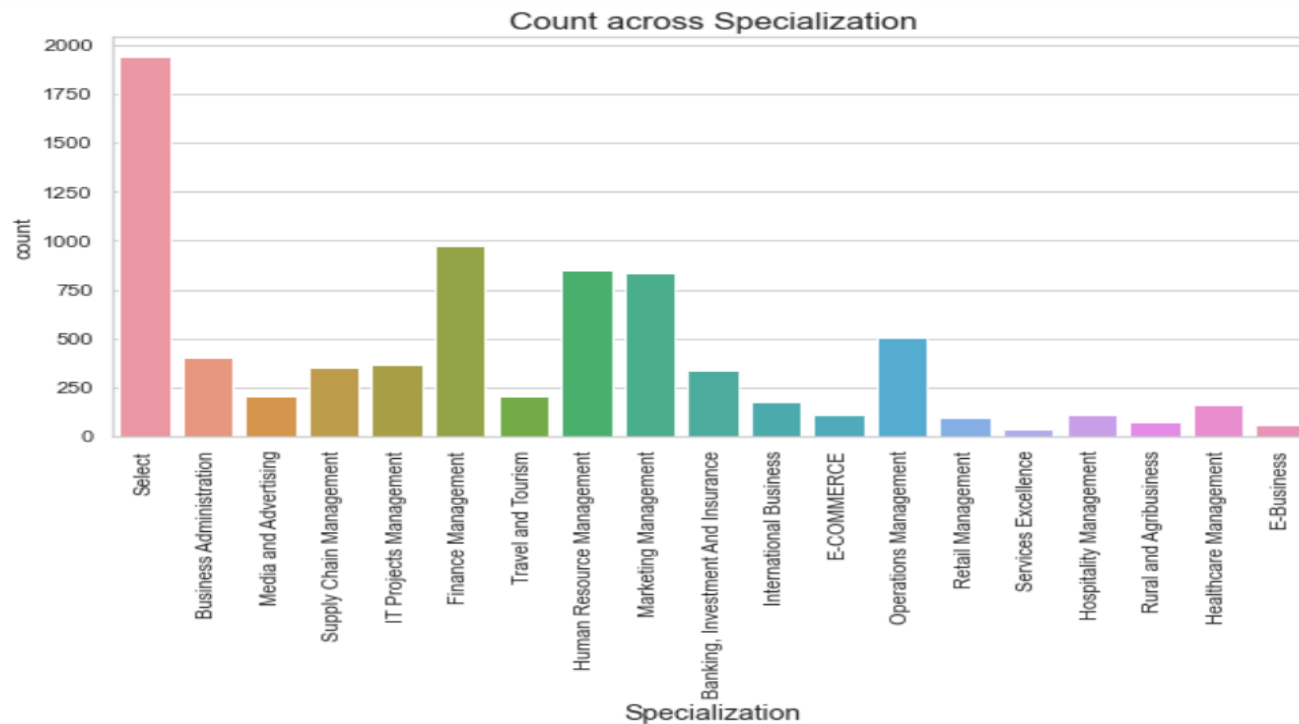
EDA



There are few columns with value 'Select'. Which means user have not selected any options. These values have been cleaned using appropriate method.

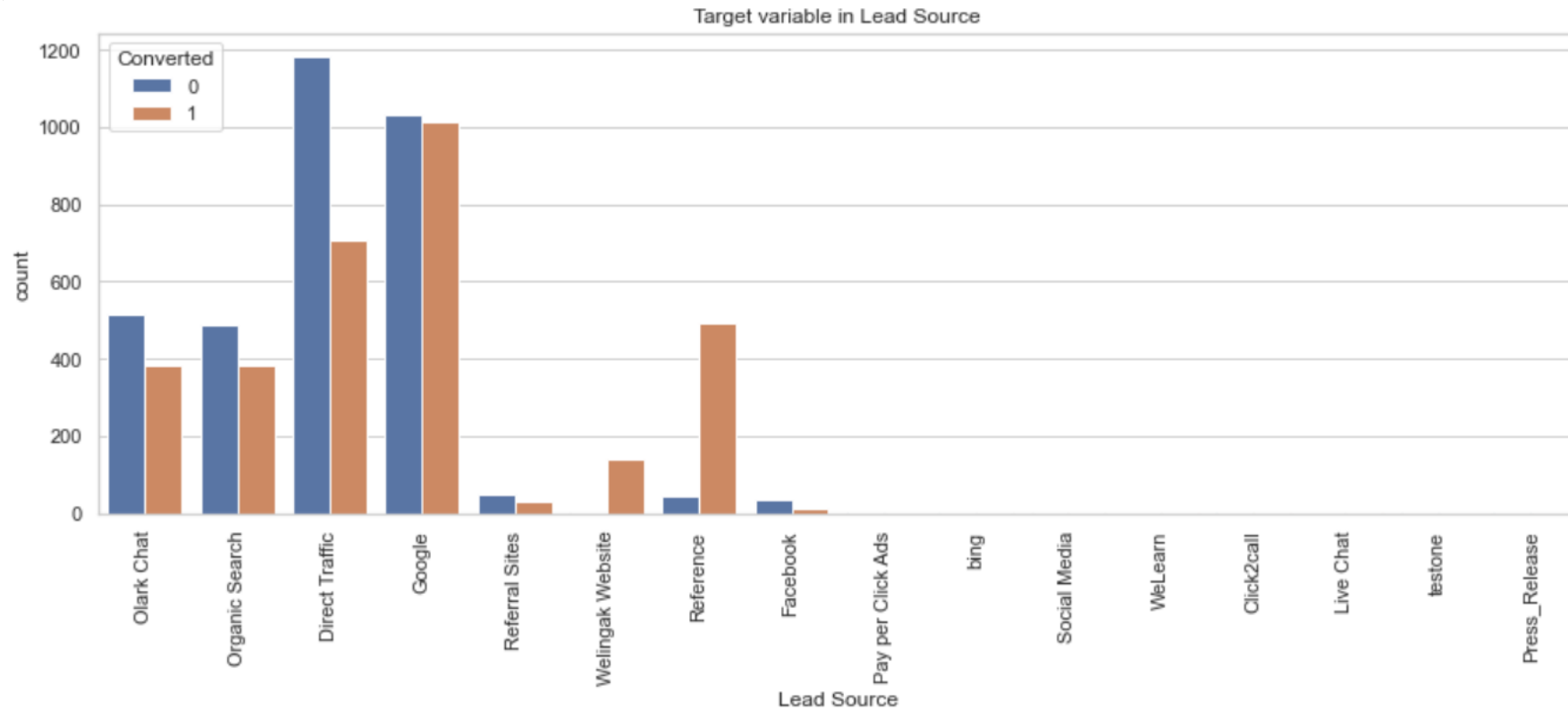
Specialization

- Leads from HR, Finance & Marketing management specializations are high probability to convert.



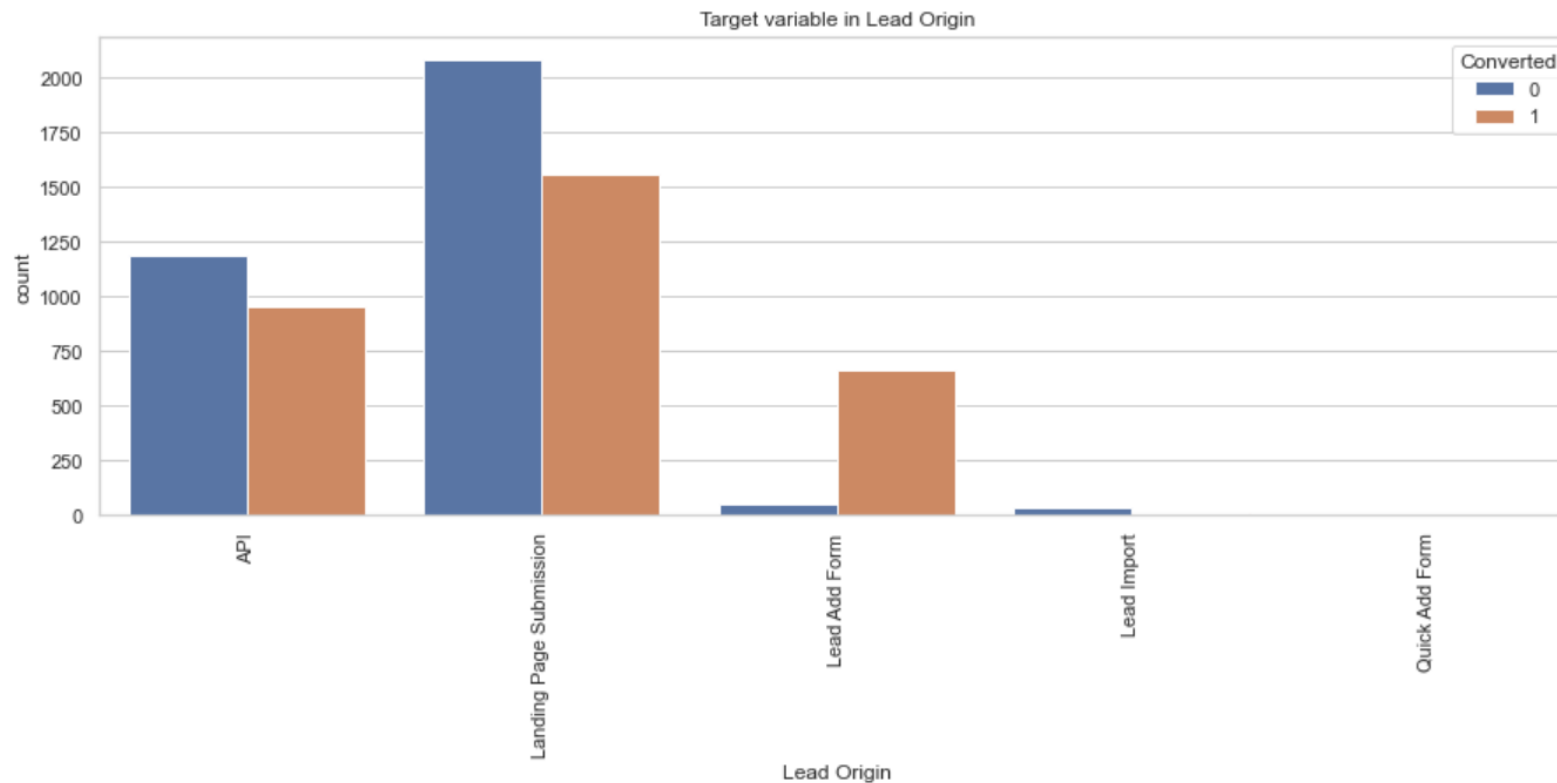
Lead Source

- In lead source the leads through google & direct traffic high probability to convert.



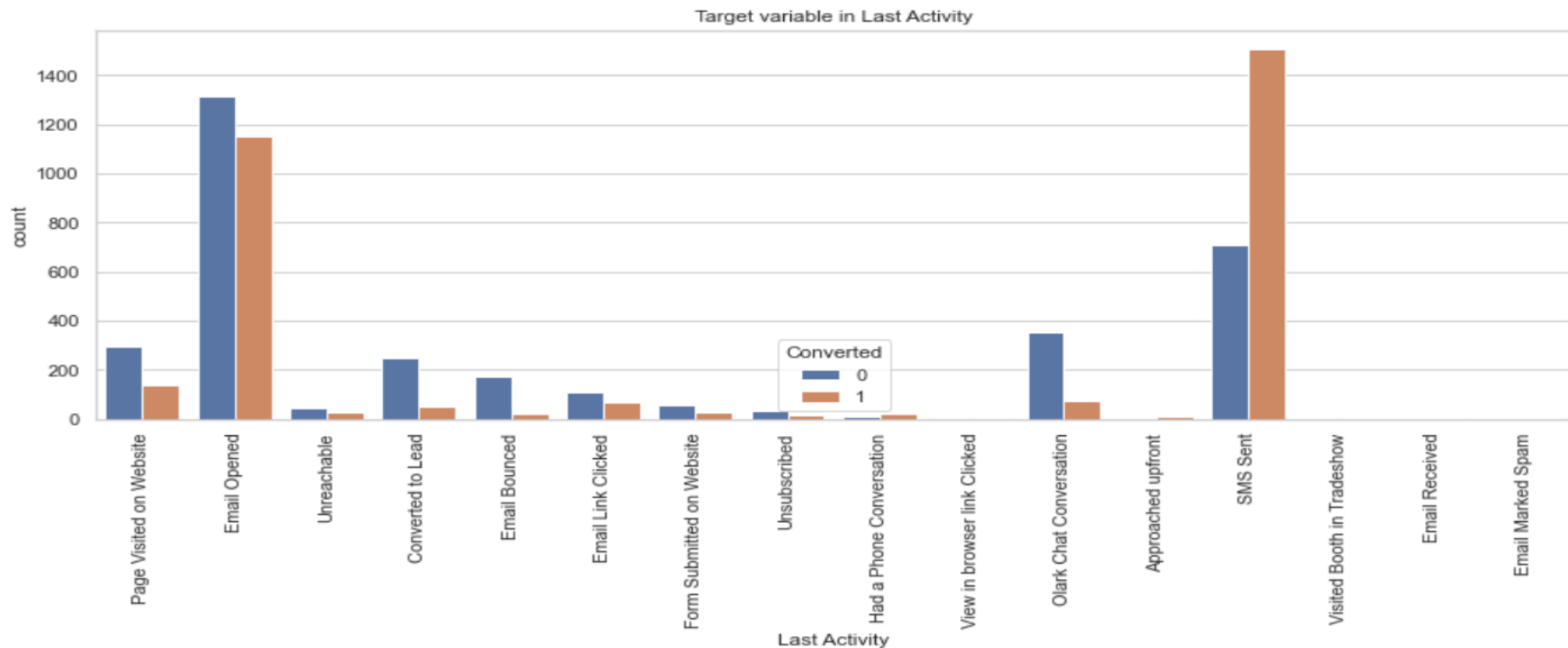
Lead Origin

- In Lead origin most number of leads are landing on submission.



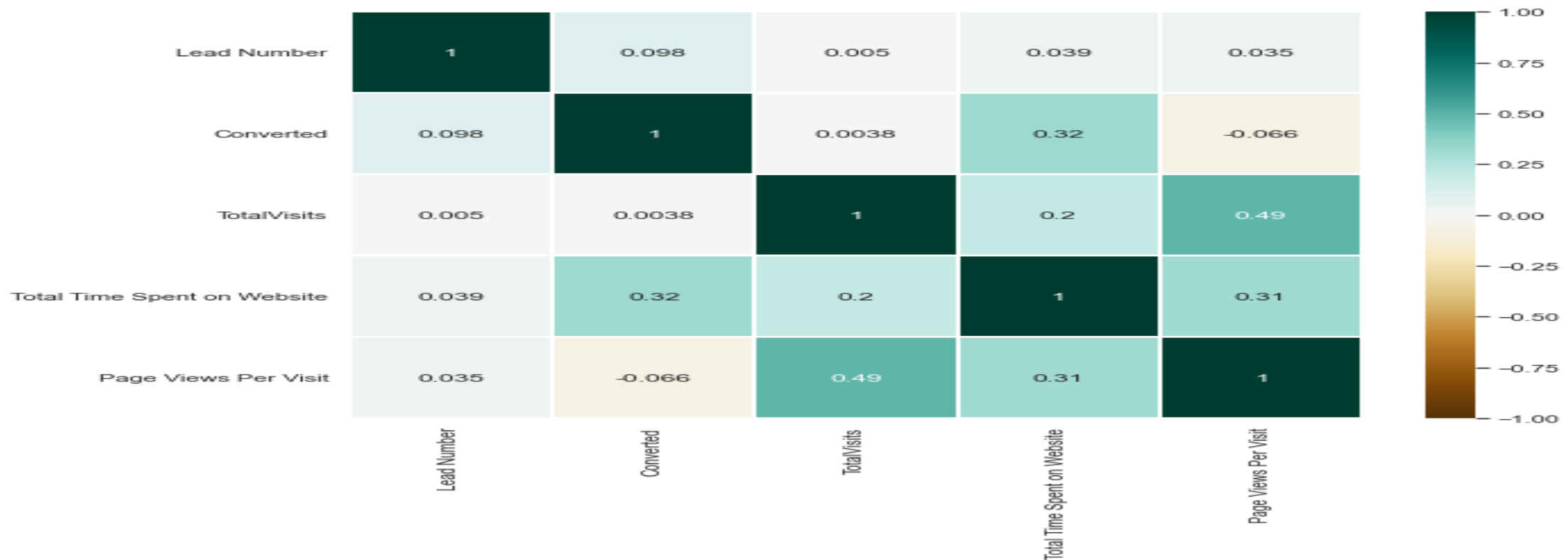
Lead Activity

- ▶ Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.



Correlation

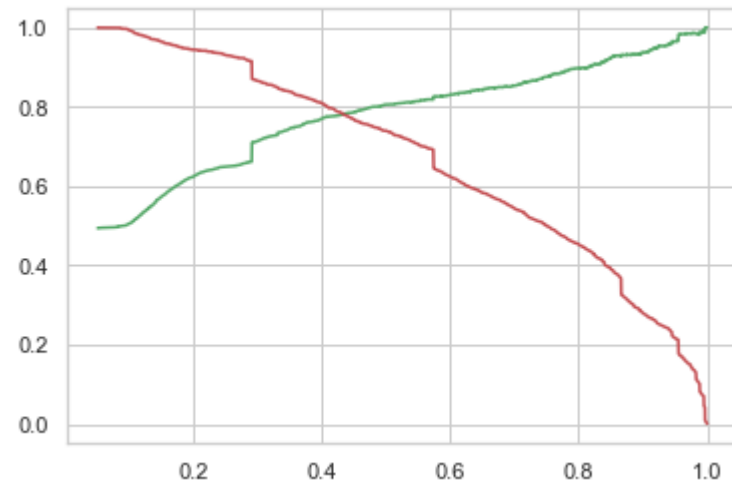
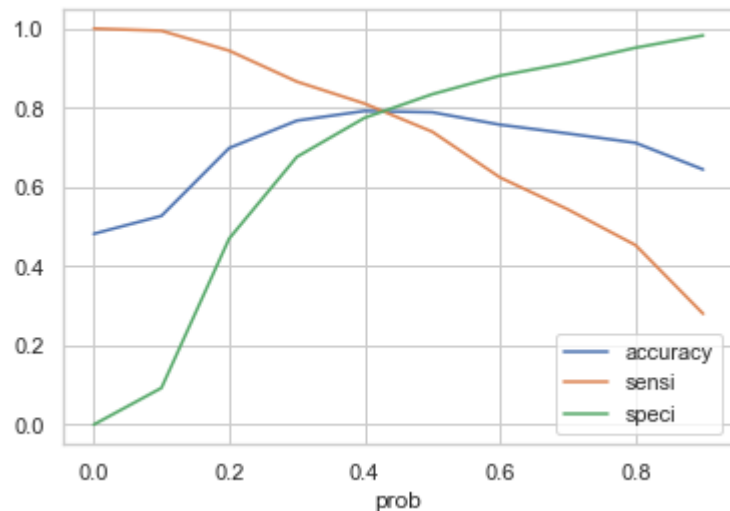
- There is no correlation between the variables.



Model Evaluation : ROC curve

► 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**



Model Parameters:

Train Data:

Accuracy : 78%
Sensitivity : 78%
Specificity : 79%

Test Data:

Accuracy : 79%
Sensitivity : 78%
Specificity : 77%

Features	Coef	P-value
const	0.2040	0.297
TotalVisits	11.1489	0.000
Total Time Spent on Website	4.4223	0.000
Lead Origin_Lead Add Form	4.2051	0.000
Lead Source_Olark Chat	1.4526	0.000
Lead Source_Welingak Website	2.1526	0.038
Do Not Email_Yes	-1.5037	0.000
Last Activity_Had a Phone Conversation	2.7552	0.001
Last Activity_SMS Sent	1.1856	0.000
What is your current occupation_Student	-2.3578	0.000
What is your current occupation_Unemployed	-2.5445	0.000
Last Notable Activity_Unreachable	2.7846	0.001

Business Results and Insights

- ▶ **Top 3 Variables Contributing to Lead Conversion:**

- ▶ **Lead Source:** Significant impact from channels like social media, referrals, and online search.
- ▶ **Total Time Spent on Website:** A strong indicator of lead interest and conversion likelihood.
- ▶ **Number of Pages Viewed:** Reflects lead engagement and is a key predictor of conversion.

- ▶ **Key Categorical Variables:**

- ▶ **Lead Source: Social Media** shows higher conversion rates.
- ▶ **Job Title: Manager/Decision Maker** indicates a higher likelihood of conversion.
- ▶ **Industry: Technology** leads are more likely to convert.

- ▶ **Strategic Recommendations:**

- ▶ **Aggressive Conversion During Intern Phase:** Prioritize high-probability leads, increase call frequency, and allocate multiple interns to key leads.
- ▶ **Minimizing Useless Calls During Slow Periods:** Raise call thresholds, focus on high-value leads, and utilize alternative communication channels.

Conclusion

- ▶ **Lead Source, Time on Site, and Pages Viewed** are the most critical factors in predicting lead conversion.
- ▶ **Focusing on social media leads, decision-makers, and tech industry** leads can significantly enhance conversion rates.
- ▶ **Adaptive strategies** for different business phases can optimize resource allocation and improve overall efficiency.
- ▶ Leads who spent more time on website, more likely to convert.
- ▶ Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.