Lead Scoring Case Study using logistic regression

SUBMITTED BY :

1. **Akshay V Doyizode**
2. **Akshita Goyel**
3. **Aishwarya Vijayakumar**

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. They have process of form filling on their website after which the company that individual as a lead.
- 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%.** Now, this means 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

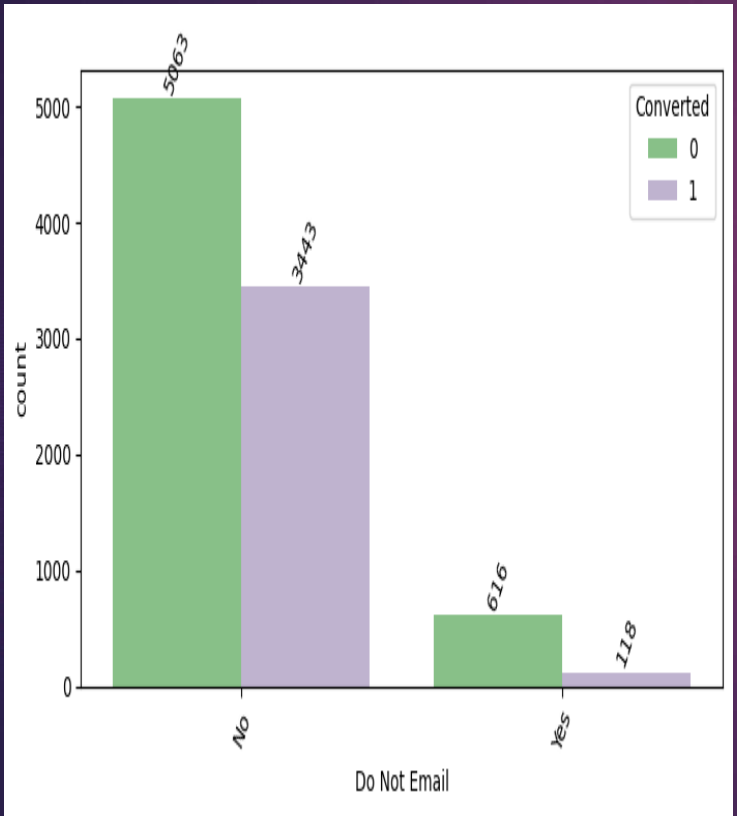
Business Objective

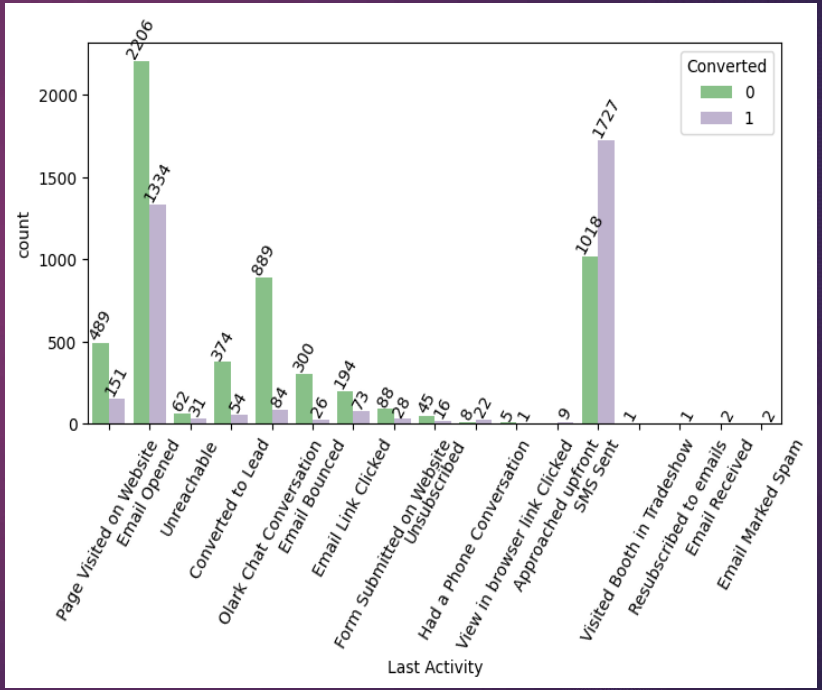
- Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can identify the Hot leads and increase their conversion rate as well.
- The CEO want to achieve a lead conversion rate of 80%.
- They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches.

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 – Data Cleaning

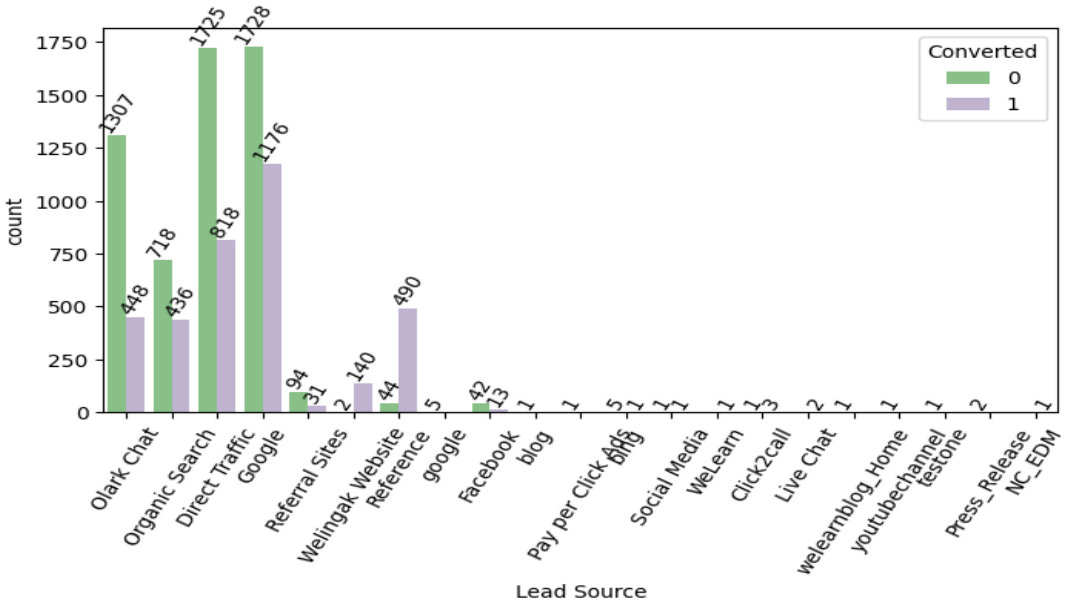


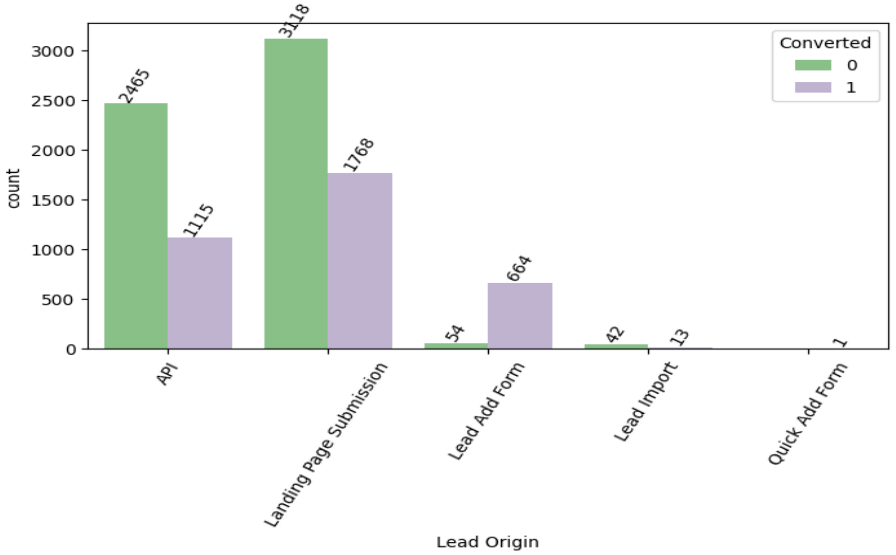


Lead Source & Lead origin

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

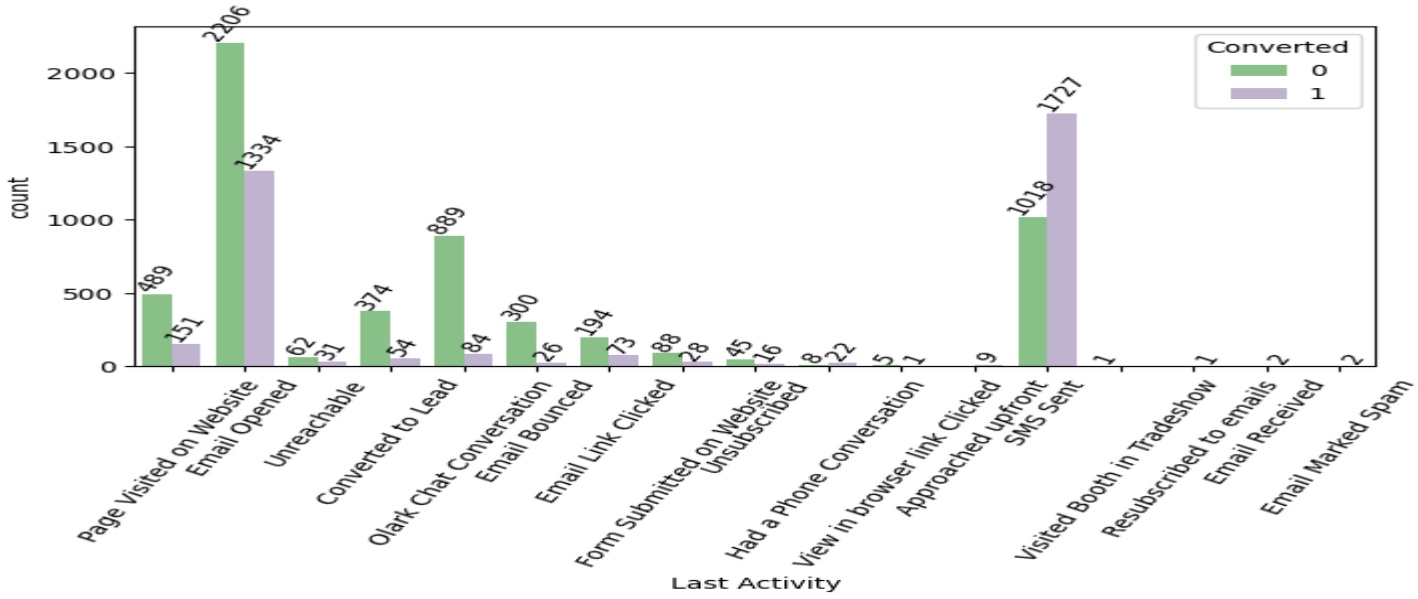
Whereas in Lead origin most number of leads are landing on submission





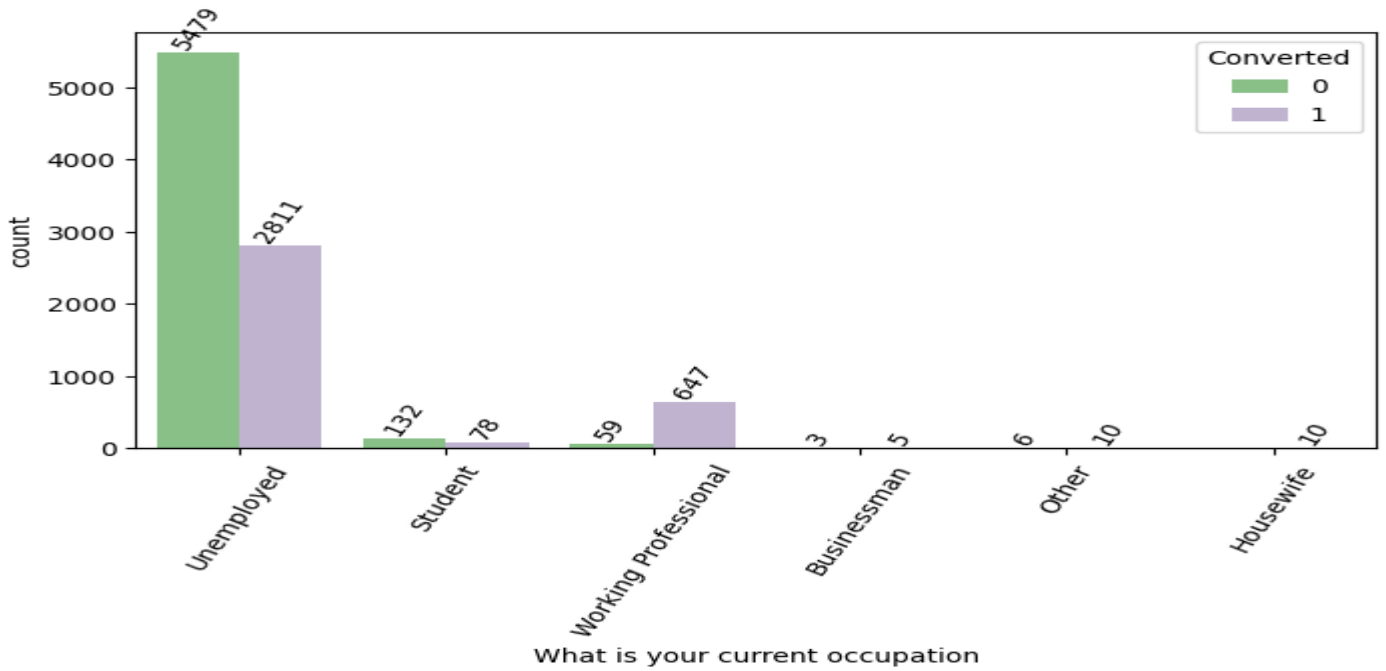
Last lead Activity

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

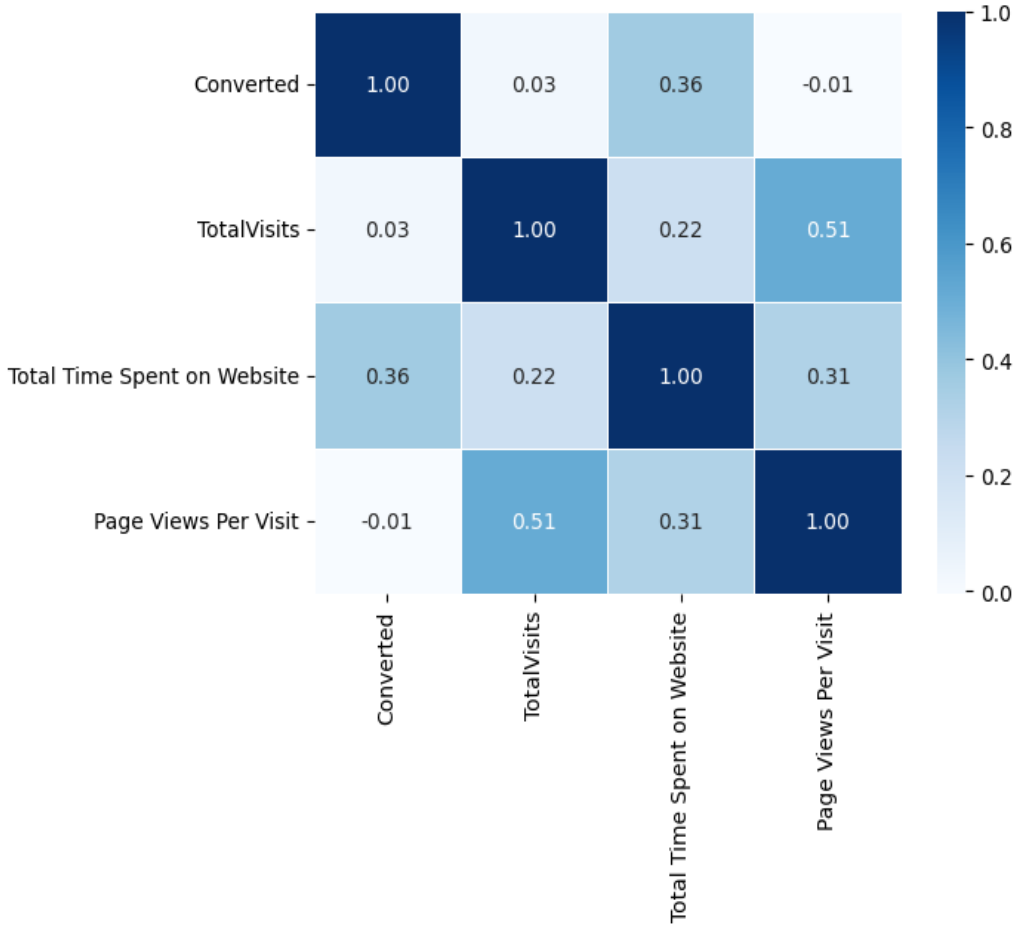


Last What is Your Occupation

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



Correlation

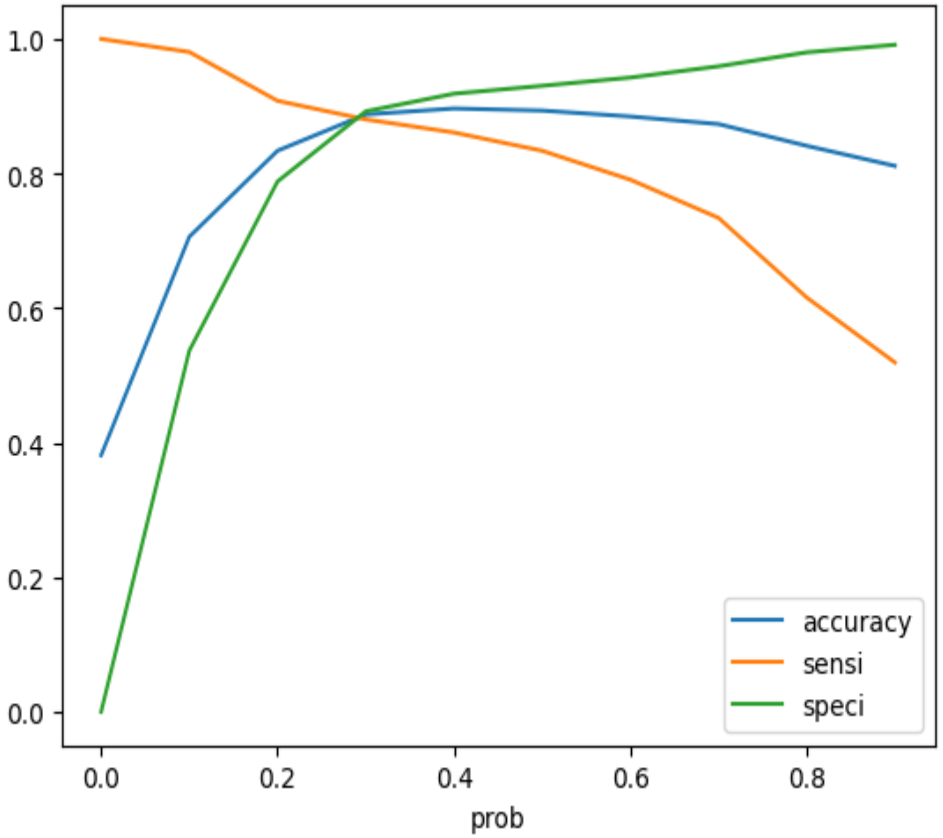


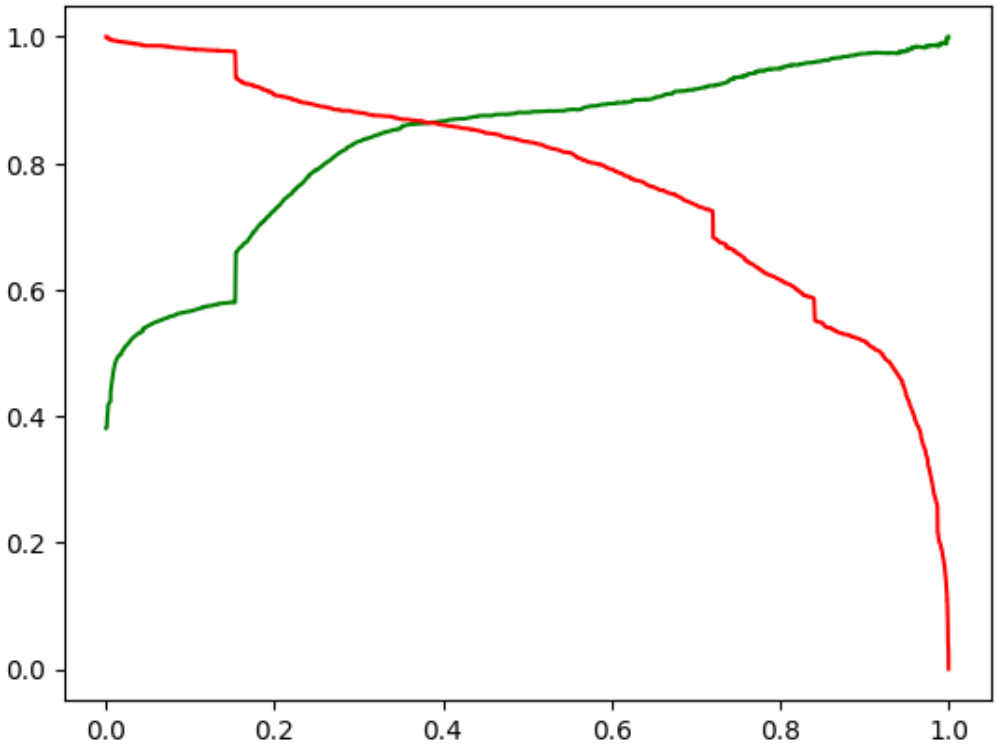
Model Evaluation

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**





Observations

Train Data:

Accuracy :89.6%
Sensitivity : 86%
Specificity : 91%

Test Data:

Accuracy:89.6%
Sensitivity : 86%
Specificity : 91%

Conclusion

1. Performance Evaluation:
   - The model exhibits strong performance metrics on both the train and test datasets, with high accuracy, sensitivity, and specificity, indicating its effectiveness in predicting conversion rates.
   - Despite a slight decrease in precision on the test dataset, the overall performance remains robust, bolstering confidence in the model's predictive capabilities.

2. Optimal Cutoff Threshold:
   - The determined optimal probability threshold of 0.42 ensures the best balance between sensitivity and specificity, contributing to accurate lead classification.

3. Key Influential Factors:
   - Various factors significantly contribute to lead conversion, including last notable activity, specific tags, lead origin, current occupation, and website engagement metrics.
   - Notable activities such as SMS sent and phone conversations, along with tags indicating interest and engagement, play pivotal roles in lead conversion.

4. Strategic Implications:
   - Leveraging identified influential factors, X_Education can strategically prioritize leads for targeted marketing efforts and personalized engagement.
   - Focus should be placed on engaging working professionals, optimizing website content and communication strategies, and leveraging insights from lead behavior to maximize conversion rates.