Lead Score Case Study

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%.

- Now, although X Education gets a lot of leads, its 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

Objective

- To select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The X Education 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 higher lead score have a higher conversion chance and the customers with 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%.
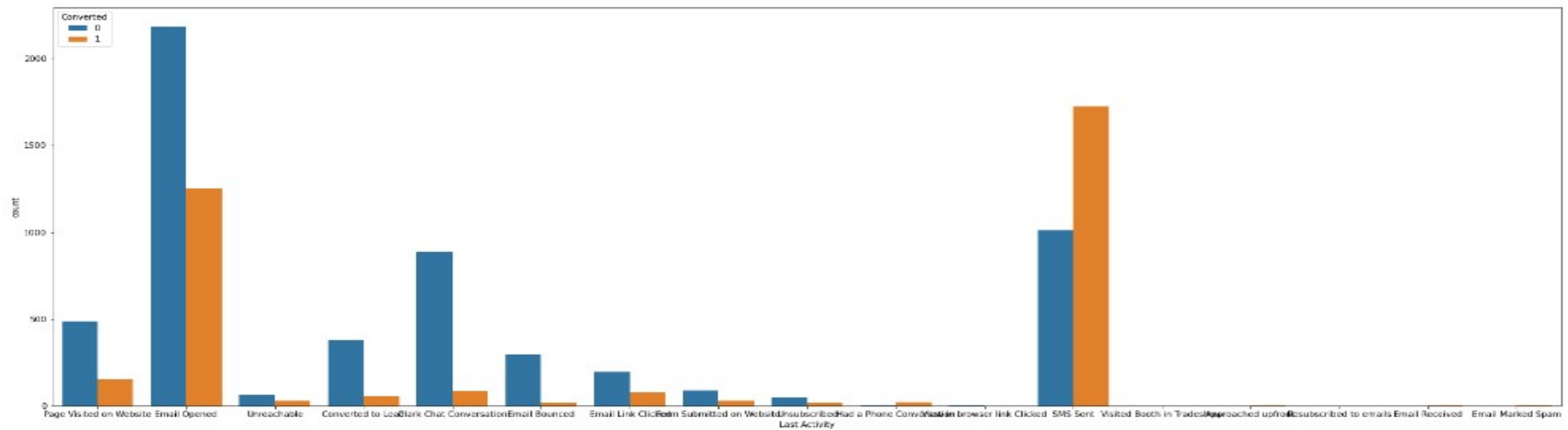
Exploratory Data Analysis on the Dataset
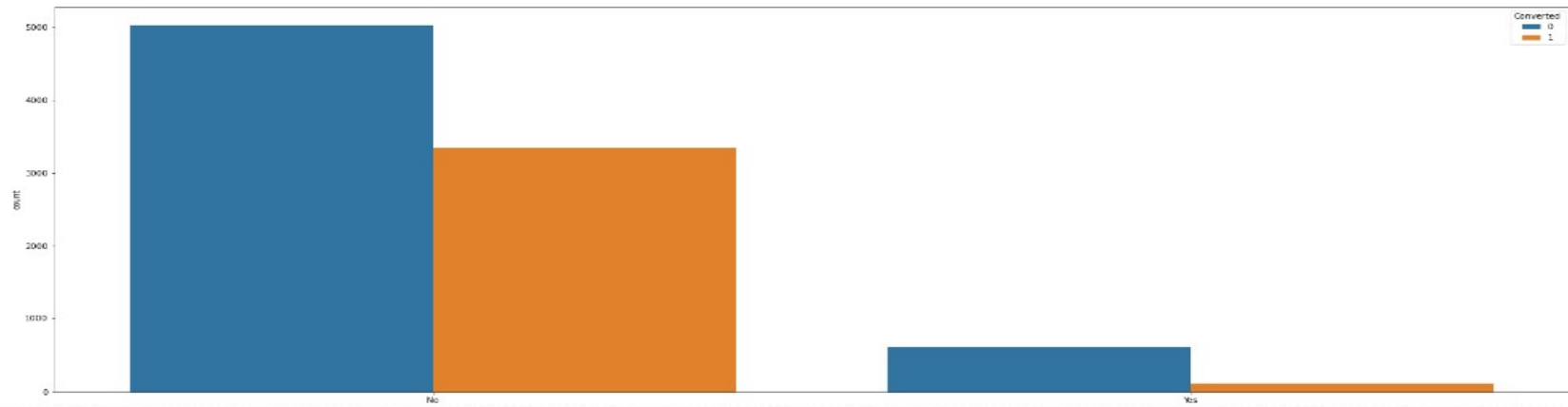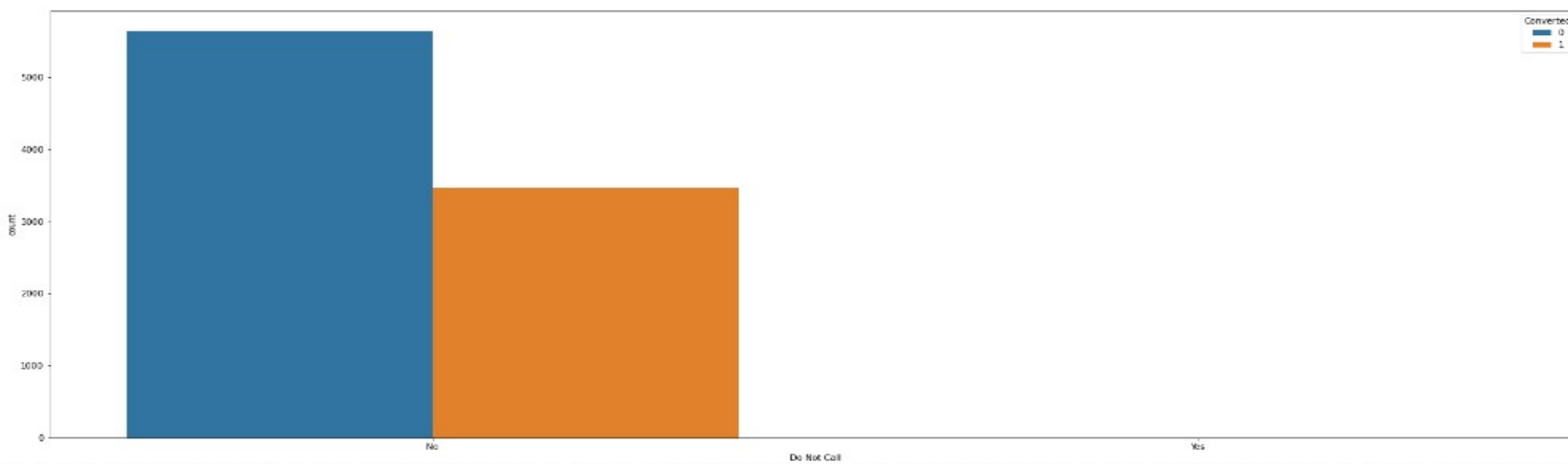
Exploratory Data Analysis on the Dataset

# Exploratory Data Analysis on the Dataset
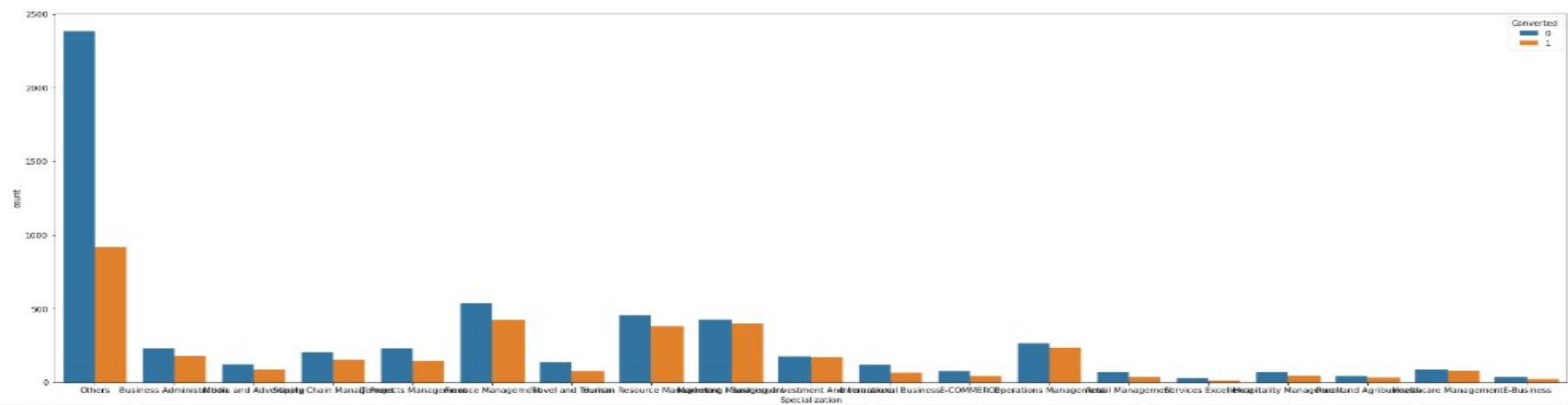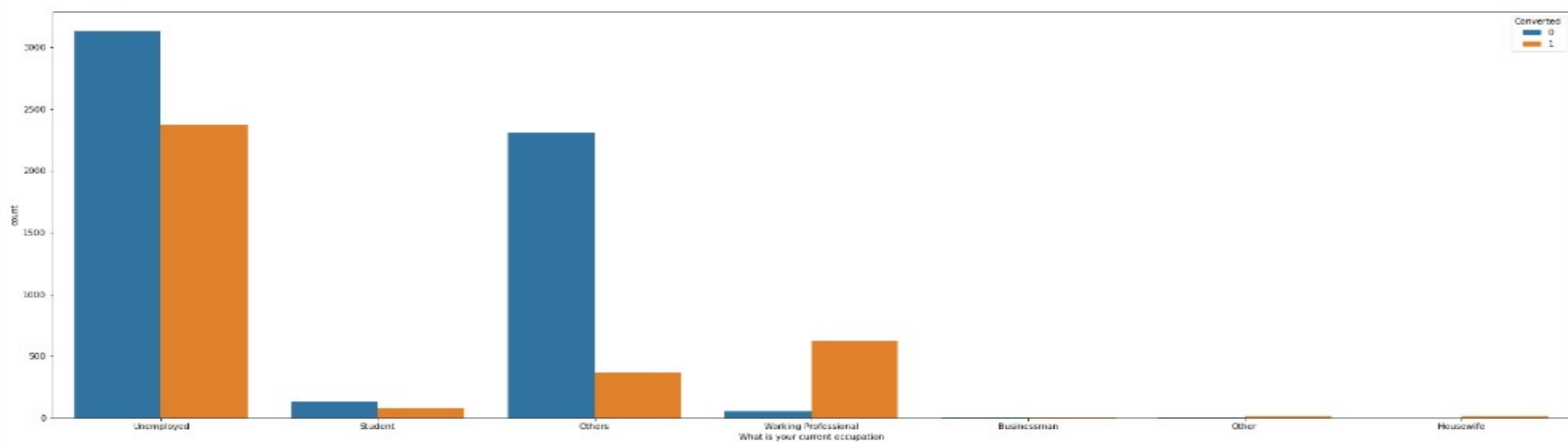
Insights from EDA on Categorical Columns

- Leads with Do not call and do not Email are not likely to take up the course
- Last Activity with Email Opened and SMS Sent are more likely to take up the course
- Majority of the leads are from India
- Most people with other or unknown specialization more likely to take up the course followed by Finance Management and Resource marketing
- Working professionals are more likely to take up the course

Correlation Between Numerical variables
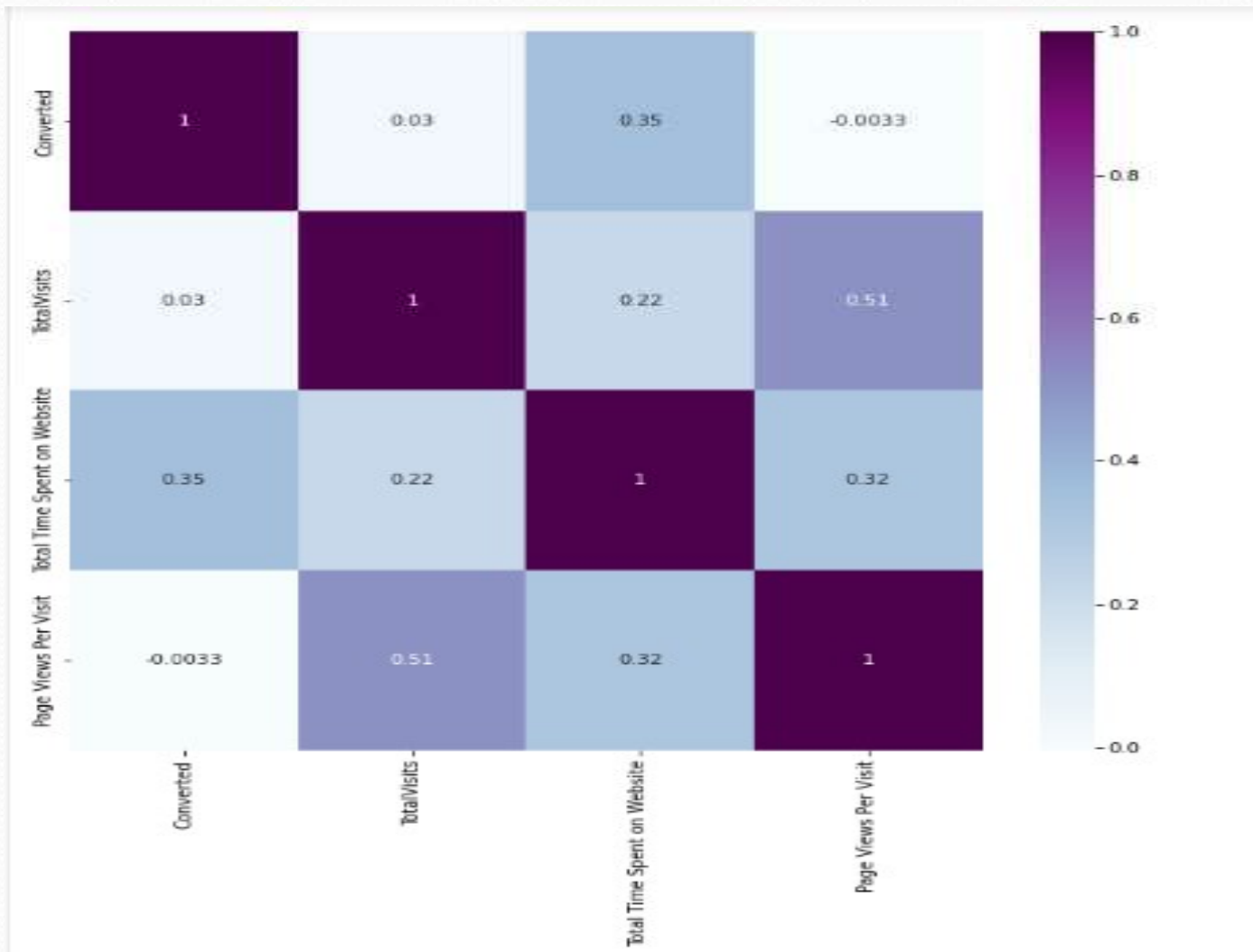
Correlation Between Numerical variables

- Total Visits and Page Views Per Visit are positively correlated with a value of 0.51
- Total Time Spent on Website has positive correlation with Page Views Per Visit with a value of 0.32

Business Suggestions

- Focus on leads who show interests like working professionals and people who want to up skill themselves

- Regular conversation and follow-up with potential leads based on their availability

- Start the next conversation based on the last activity/conversations with the leads

- Do proper research about the interests of the leads before the conversation based on their application details and tell them the benefits they will get after they pursue their course from X Education company

Business Suggestions

- Instead of making regular calls, write a detailed email or message about the courses and the benefits they get to the less potential lead, based on their interests they may call back.

- Conduct a master class about the session instead of making phone calls to individuals that will save a lot of time as well as increase the reach to mass audience.

- Also display the Alumni Success status and feedbacks on the website by signing up for X Certifications from X Education Company.

ML Model Predictions

The top three variables in the model which contribute most towards the probability of a lead getting converted are:

- **Tags_Lost to EINS** with coefficient of 9.7293
- **Tags_Closed by Horizzon** with coefficient of 8.4774
- **Tags_Will revert after reading the email** with coefficient of 6.7125

The top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are :

- **Tags_Will revert after reading the email**
- **Last Notable Activity_SMS Sent**
- **Tags_Busy**

Optimal cut-off
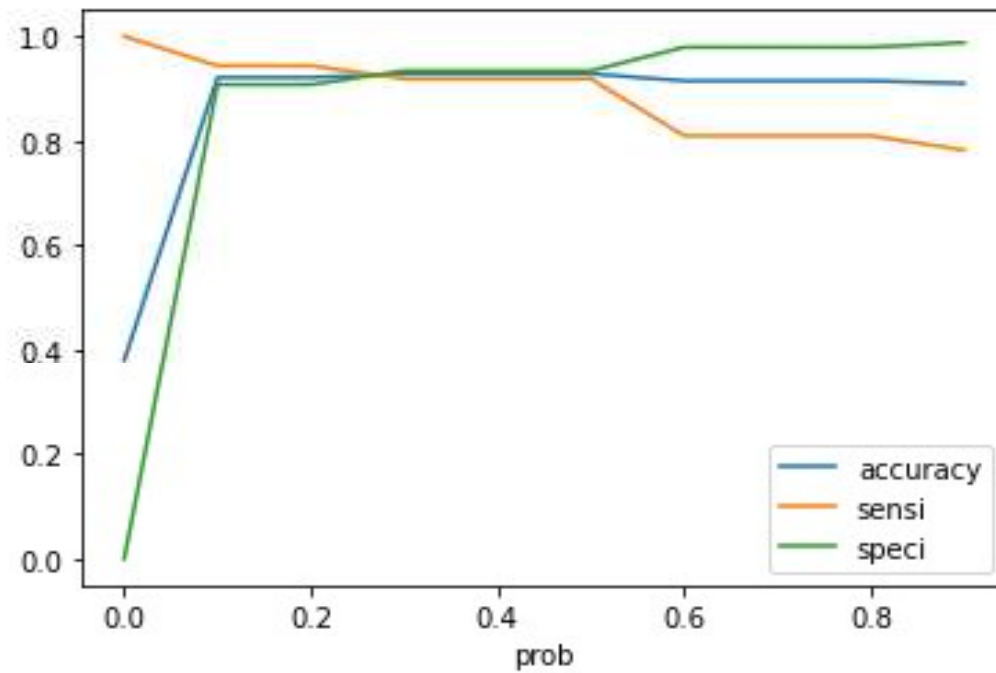
From the curve above, 0.3 is the optimum point to take it as a cutoff probability.

Model Evaluation

Evaluation Metrics on the Train dataset

- Sensitivity : 91.85 %
- Specificity : 93.37 %
- False Positive rate : 6.62 %
- Positive Predictive value : 89.45 %
- Negative Predictive value : 94.93 %

Model Evaluation

Evaluation Metrics on the Test dataset

- Sensitivity : 90.49 %
- Specificity : 93.84 %

Thank you

Submitted by

Rajkumar B and Hanumanth A