LEAD SCORE CASE STUDY

SUBMITTED BY : ANIKET TRIPATHI

Problem Statement: X Education is an online course provider that specializes in selling courses to professionals in various industries. The company promotes its courses on different websites and search engines like Google. When visitors arrive on the website, they have the option to explore the available courses, complete a form to express interest, or watch informational videos. By providing their email address or phone number, these individuals are considered leads. Additionally, X Education also receives leads through referrals from past customers.

Once these leads are obtained, the sales team initiates contact by making phone calls, sending emails, and employing other communication methods. During this process, some leads are successfully converted into paying customers, while the majority do not proceed further. On average, X Education achieves a lead conversion rate of approximately 30%.

Business Goal: X Education is seeking assistance in identifying the most prospective leads, those with the highest probability of becoming paying customers. They require a model that assigns a lead score to each lead, indicating the likelihood of conversion. Customers with higher lead scores are expected to have a greater chance of converting, while those with lower scores are expected to have a lower likelihood of conversion. The CEO has provided a rough target lead conversion rate of approximately 80%.

Strategy

- Load the data for analysis

- Clean the data and prepare for analysis

- Exploratory Data Analysis.

- Feature Scaling

- Splitting the data into Test and Train dataset.

- Building a logistic Regression model and calculate Lead Score.

- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.

- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

Problem Solving Methodology

Data Sourcing , Cleaning and Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization.
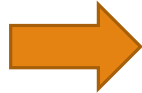
Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Splitting data into train and test set.

Model Building

- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.
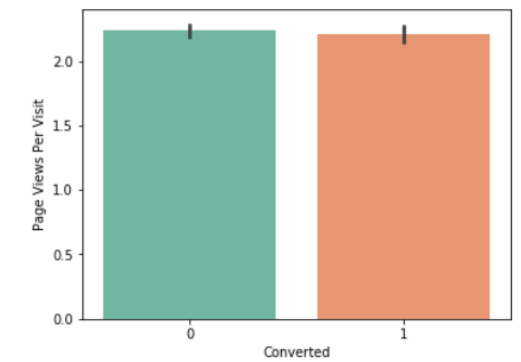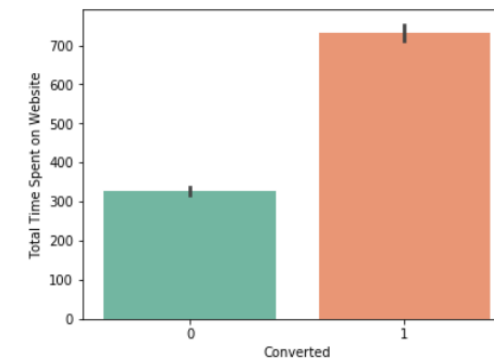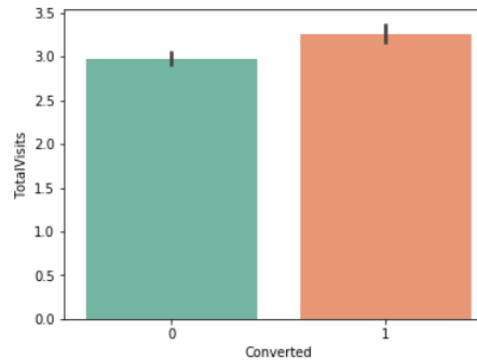
Result

- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics
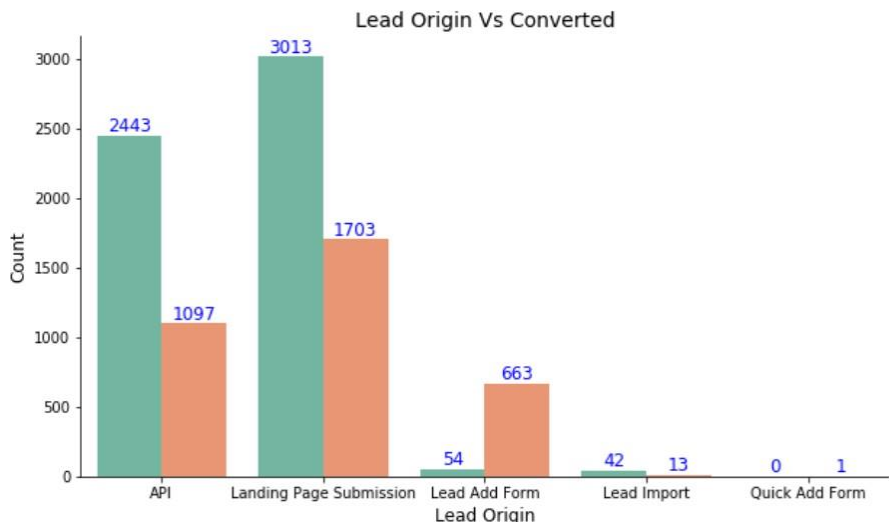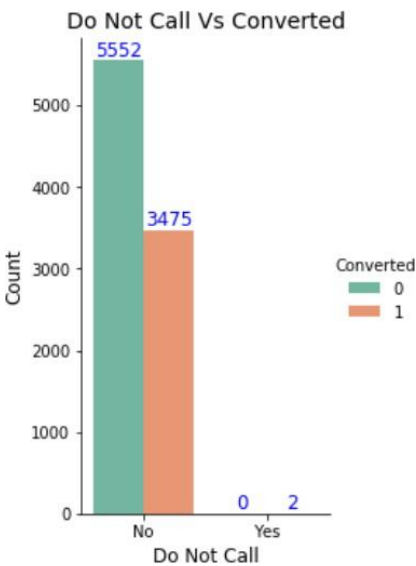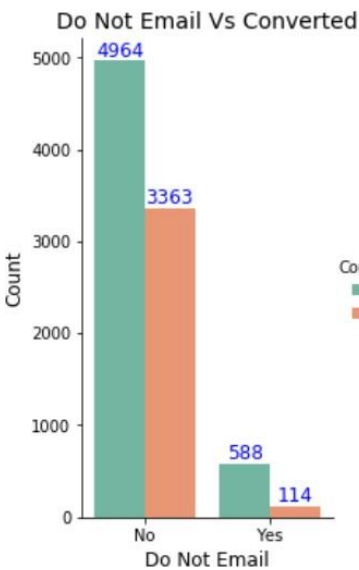
We have around 39% Conversion rate in Total
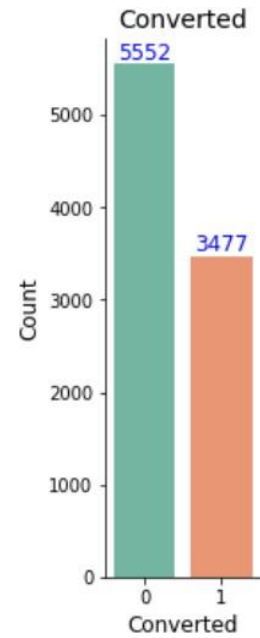
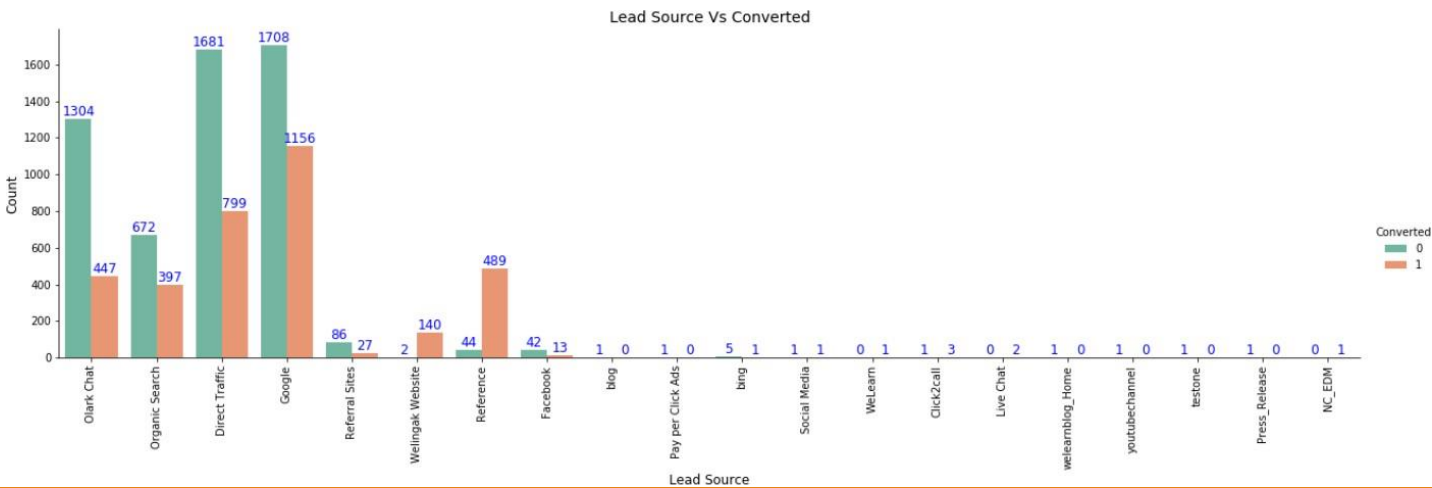The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit

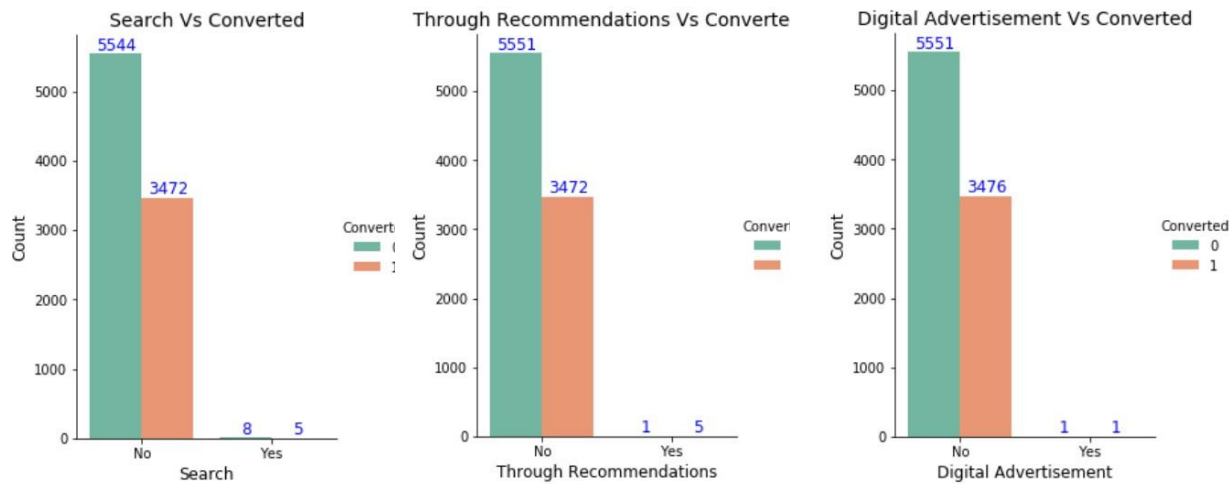In Lead Origin, maximum conversion happened from Landing Page Submission

Major conversion has happened from Emails sent and Calls made

Lead Origin Vs Converted

Do Not Email Vs Converted

Do Not Call Vs Converted

Major conversion in the lead source is from Google
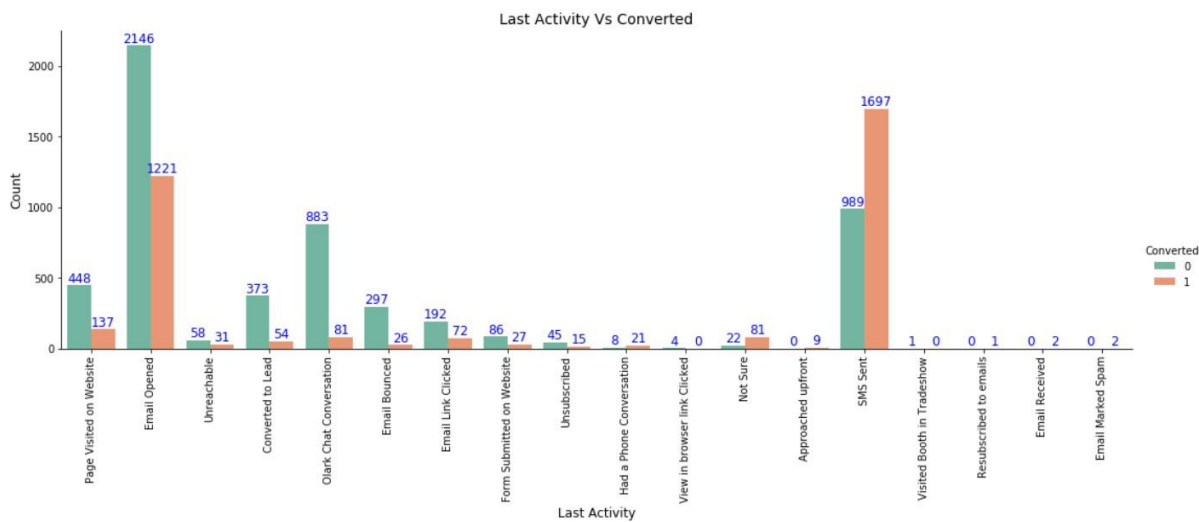

Lead Source Vs Converted

Not much impact on conversion rates through Search, digital advertisements and through recommendations

More conversion happened with people who are unemployed
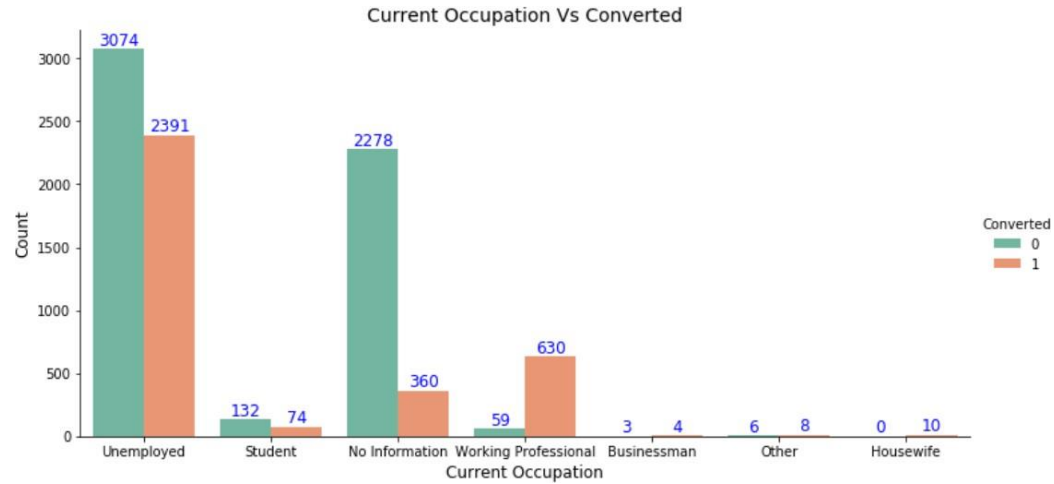
Last Activity value of SMS Sent' had more conversion.

Variables Impacting the Conversion Rate

- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

## Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.37 based on Accuracy,
Sensitivity and Specificity
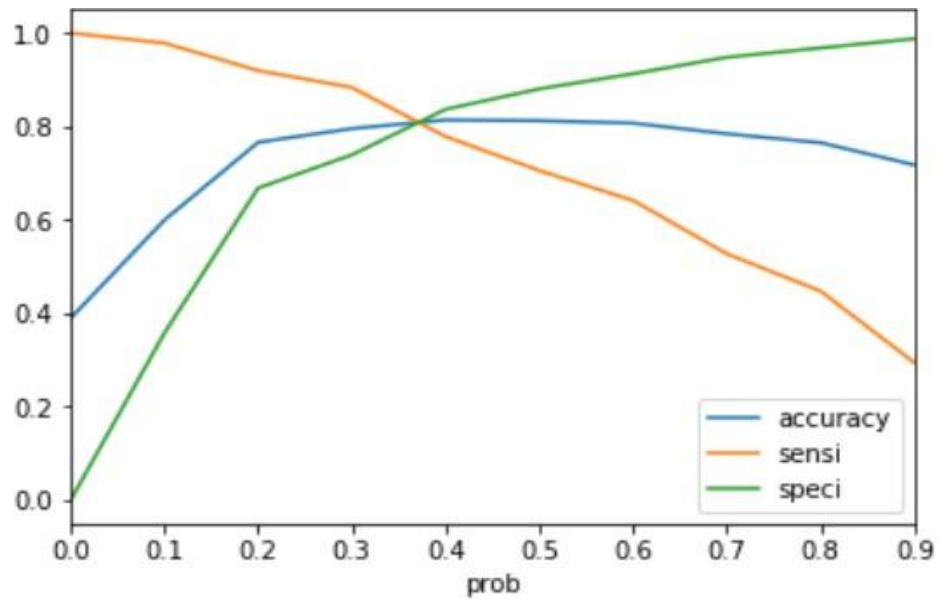
- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%

Model Evaluation- Precision and Recall on Train Dataset

The graph depicts an optimal cut off of 0.42 based on Precision and Recall
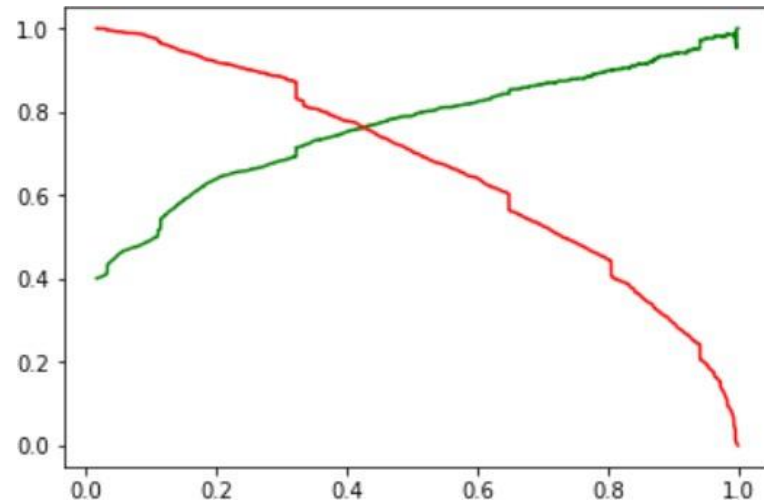
- Precision - 79 %
- Recall - 71 %

Conclusion

X Education is seeking assistance in identifying the most prospective leads, those with the highest probability of becoming paying customers. They require a model that assigns a lead score to each lead, indicating the likelihood of conversion. Customers with higher lead scores are expected to have a greater chance of converting, while those with lower scores are expected to have a lower likelihood of conversion. The CEO has provided a rough target lead conversion rate of approximately 80%.

The top 3 variables that contribute for lead getting converted in the model are
- Total time spent on website
- Lead Add Form from Lead Origin
- Had a Phone Conversation from Last Notable Activity

Overall this model seems to be good.