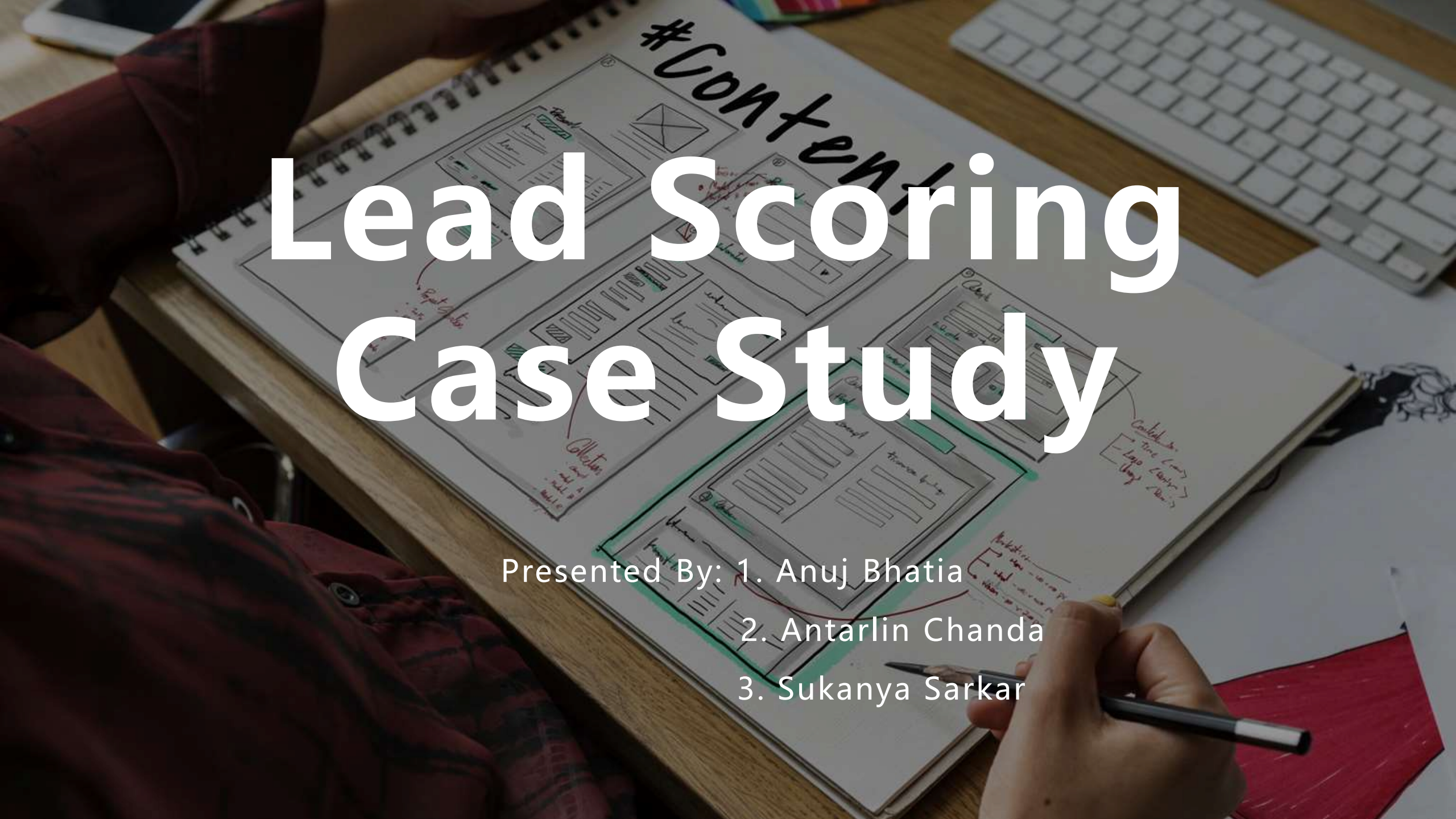


Lead Scoring Case Study

Presented By: 1. Anuj Bhatia
2. Antarlin Chanda
3. Sukanya Sarkar

Problem Statement

- X Education offers online courses for professionals.
- They market these courses on platforms like Google.
- Leads are generated through website visits, form submissions, and referrals.
- The sales team converts these leads into customers.
- The typical lead conversion rate is approximately 30%, a key performance metric.

Business Goal:

•X Education wants a lead scoring model to identify high-conversion probability leads.

•Key steps:

- Gather historical lead data.
- Select important lead factors (source, interactions, demographics).
- Train a machine learning model (e.g., logistic regression).
- Set a conversion threshold (CEO's target: 80%).
- Apply the model to new leads to assign scores.
- Adjust the score threshold to meet the target.

•Continuously monitor and improve the model.

•Implement it in lead management for prioritizing sales efforts.

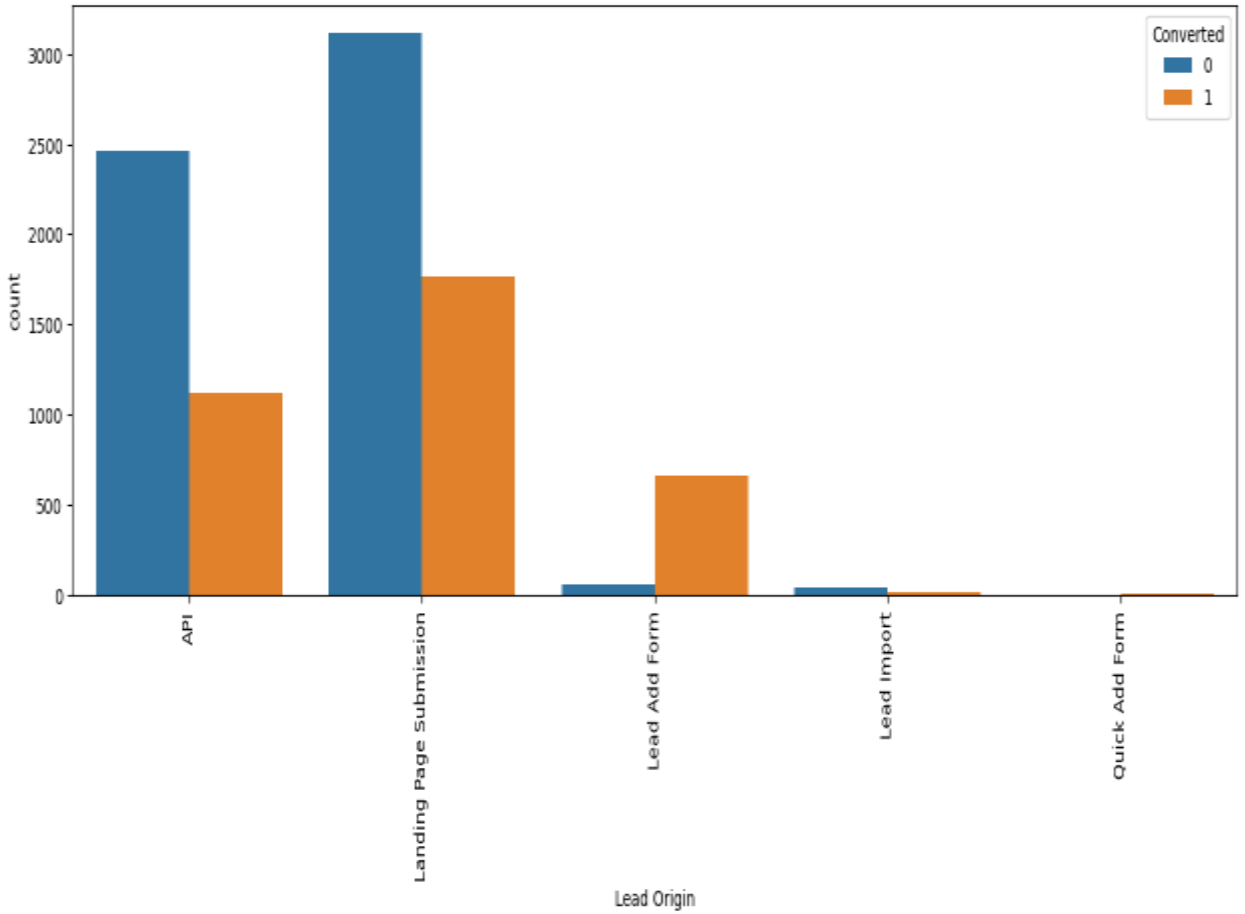
Strategy

1. **Data Collection**: Gather lead data.
2. **Data Prep**: Clean and preprocess the data.
3. **EDA**: Explore the data.
4. **Feature Scaling**: Adjust feature scales if necessary.
5. **Data Splitting**: Divide data into training and testing sets.
6. **Model Build**: Create a logistic regression model.
7. **Model Evaluation**: Assess performance with metrics like Sensitivity and Specificity.
8. **Model Application**: Use the best model on test data.
9. **Iterate and Refine**: Improve the model if needed.
10. **Deployment**: Integrate the model for automated lead scoring.
11. **Monitoring**: Continuously watch model performance and update it with new data.

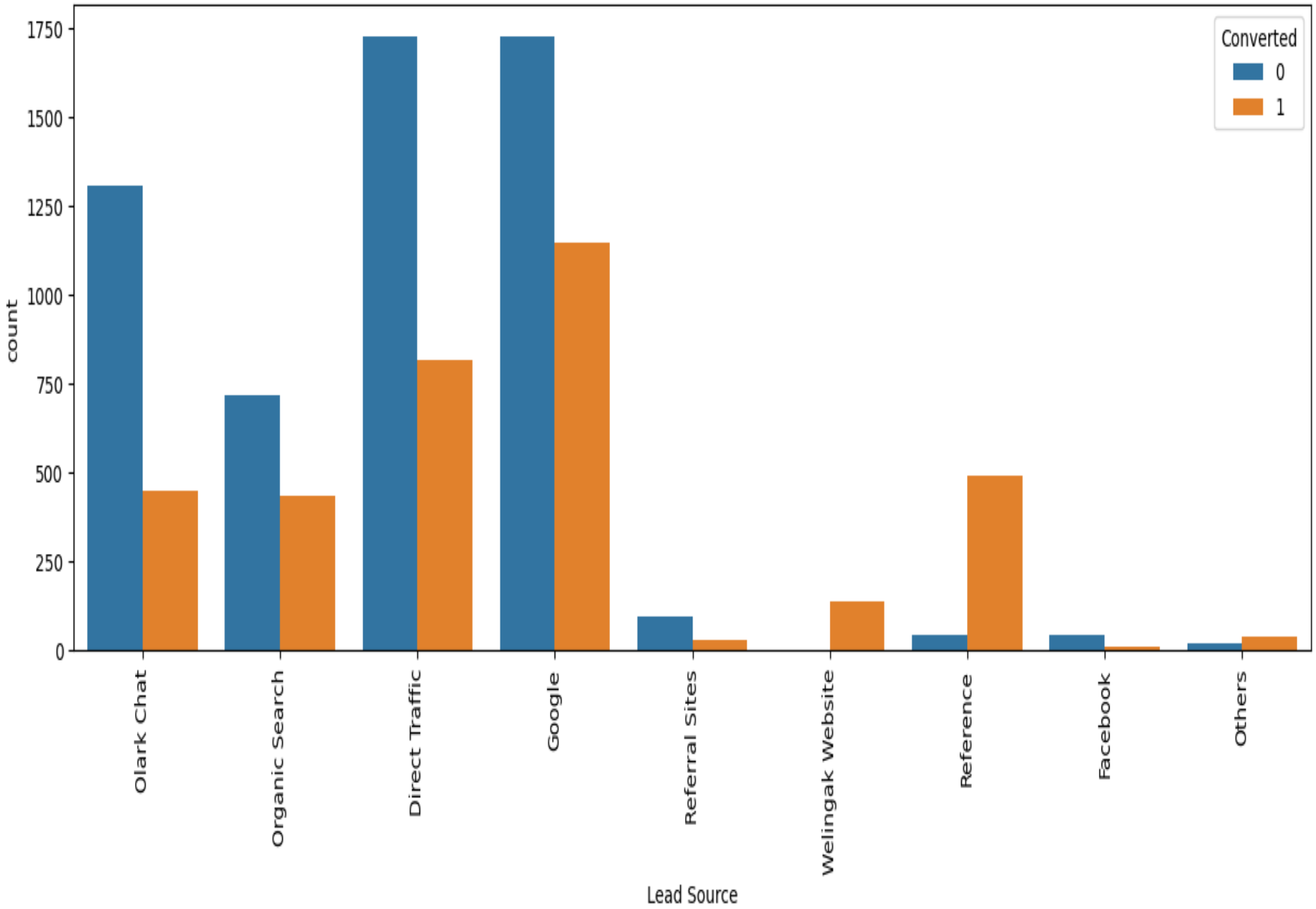
Problem Solving Methodology

- Data Sourcing , Cleaning and Preparation
- Feature Scaling and Splitting Train and Test split
- Model Building
- Result

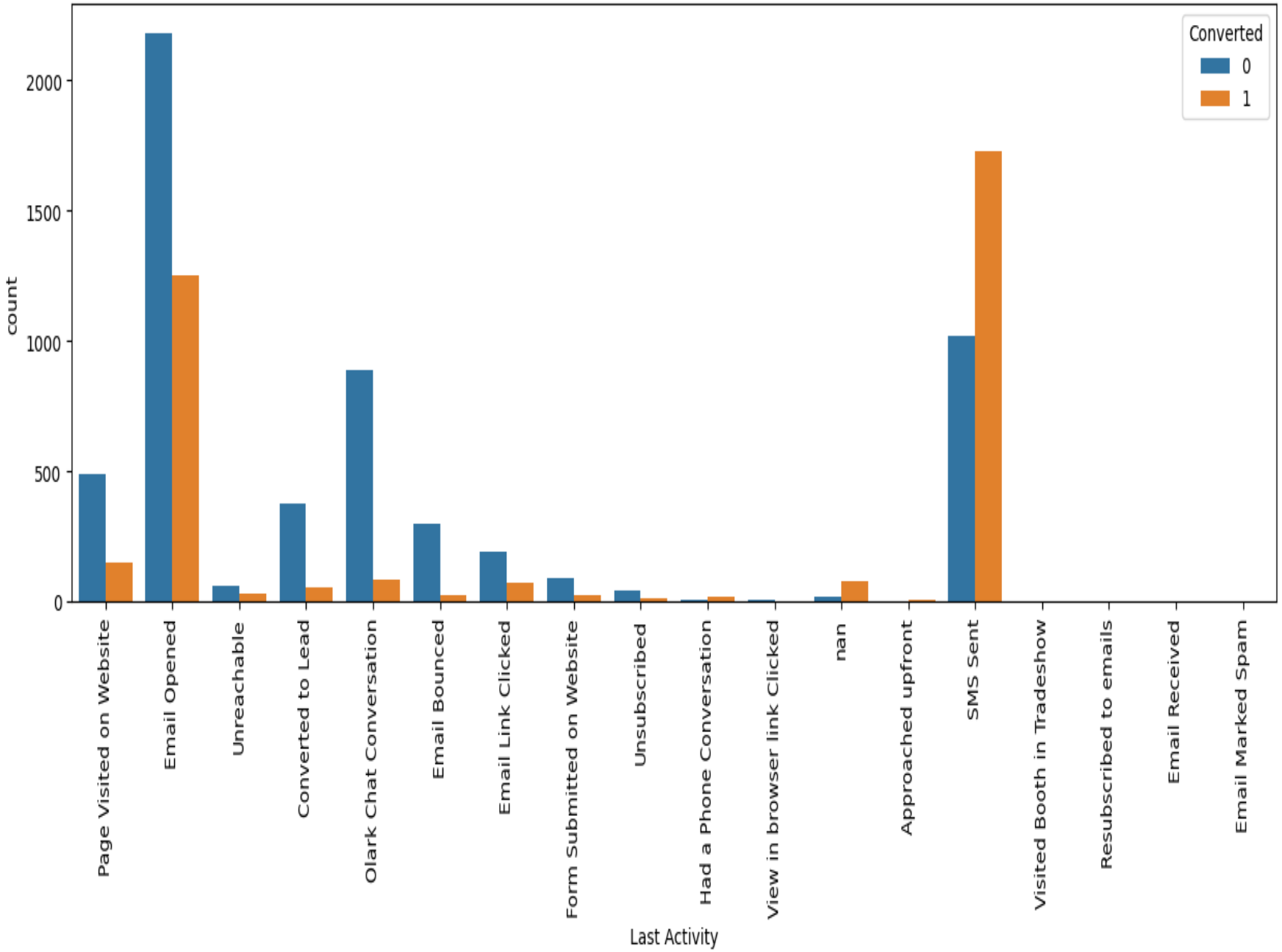
EDA (Exploratory Data Analysis)



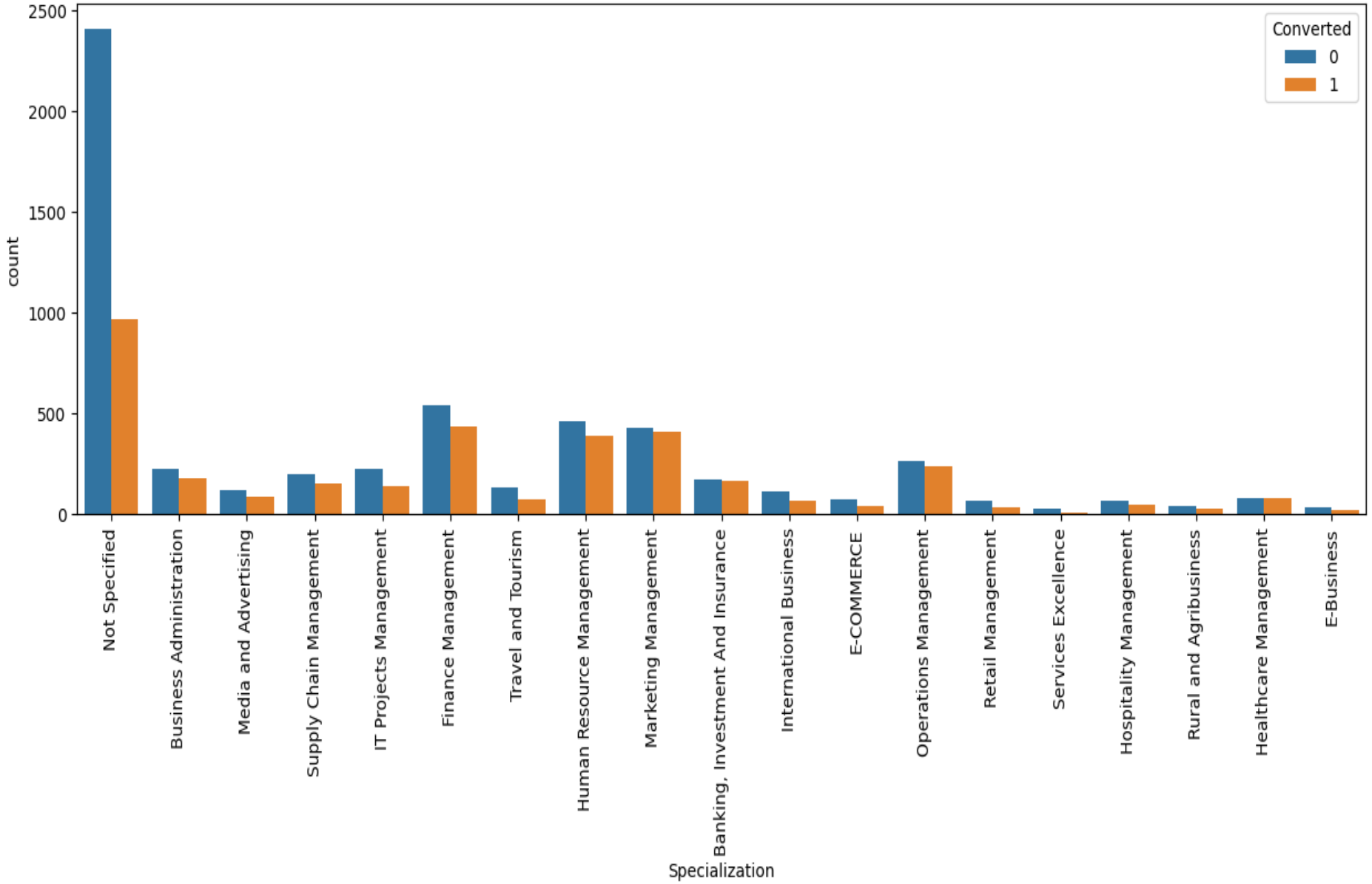
We found most leads originated from API and Landing Form Submission. Lead Add Form is the highest Yielding in terms of conversion rates.



We find the top Lead Origin Sources , Google and Reference to be high yielding for conversions



SMS Sent, and Email Opened are conscious choices so these perhaps are more worth pursuing for leads.

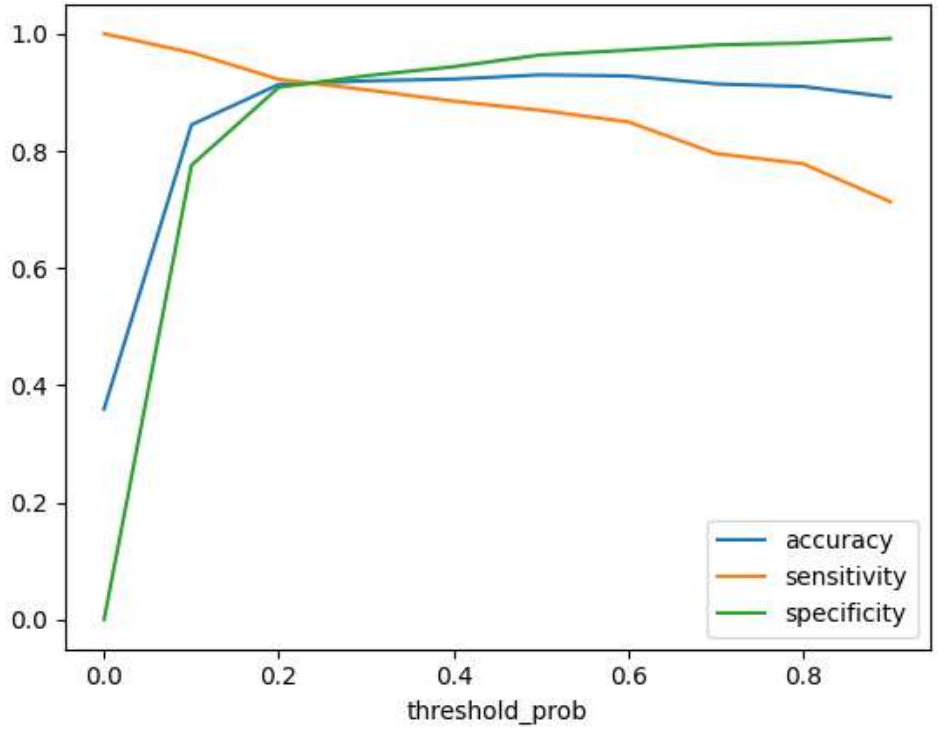


We can see HR, Finance, Marketing have good conversion rates.

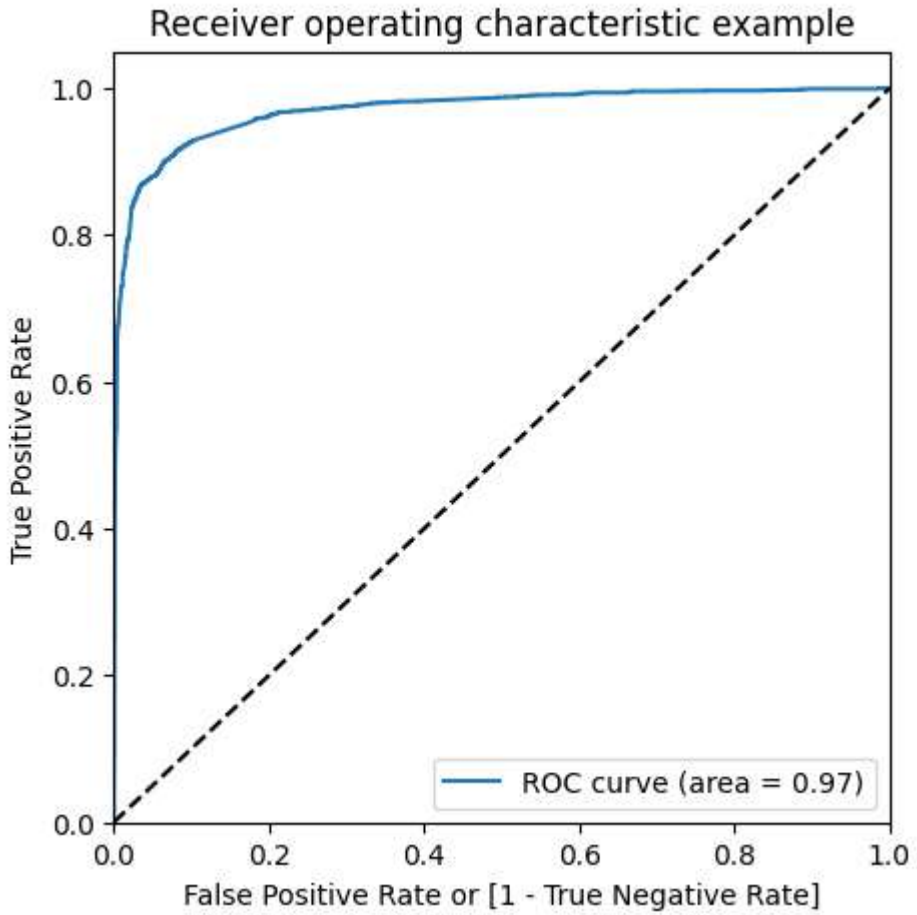
Variables Impacting the Conversion Rate

- 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



- The intersection among the metrics is the ideal point to choose for deciding the threshold since this ensures the combined maximization of all 3 entities.
- As observed the value is around 0.3.
- This is also suitable and almost in accordance with business CEO's requirement of an 80 percent conversion rate, So ideally only 20 percent of people are not potential leads.
- We got a close value of 30 percent of people who may not be that important to business.
- We can thus focus on the remaining 70 percent of people classified as converted and thus formulate a strategy for people showing similar characteristics as this 70



The ROC Curve should be a value close to 1. Our Predictive Model is good since we are getting almost 0.97 R-squared

Conclusion

(Train-Test Performance Comparison)

Train Performance:

- Accuracy: 91.95%
- Sensitivity: 86.91%
- Specificity: 96.38%
- Precision: 87.59%
- Recall: 90.41%

Test Performance:

- Accuracy: 91.45%
- Sensitivity: 90.74%
- Specificity: 91.86%
- Precision: 86.51%
- Recall: 90.74%

Insights found during EDA and Model Prediction

- People spending higher than average time are promising leads, So targeting them can be helpful for conversion.
- More advertisements in google since high yielding and also referral bonuses to past converted leads so that they recommend more people
- Working Class people, MBA Professionals or students can be focussed on for more conversion. Even specialized courses can be designed for MBA professionals.
- SMS and Emails can be improved since they are very important factors in ensuring conversion. So proper framing and design can be focussed on.
- References and offers for referring a lead can be a good source for Conversion.
- An alert messages or information has seen to have high lead conversion rate.

Final Insights

•We can see above that train and test sets have similar performances and thus there is no significant overfitting or performance loss. we can assume model can be relied on for real world data now.

•Since threshold was found to be around 0.3 so all persons above that or lead score 30 should be in our priority list and customized mails,ads and services and discounts can be offered to them to ensure conversion from them.

• Top 3 Variables for lead conversion are:

a. Tags_Will revert after reading the email
b. Tags_Closed by Horizzon
c. Tags_Ringing

•Overall the performance of the model is satisfactory.