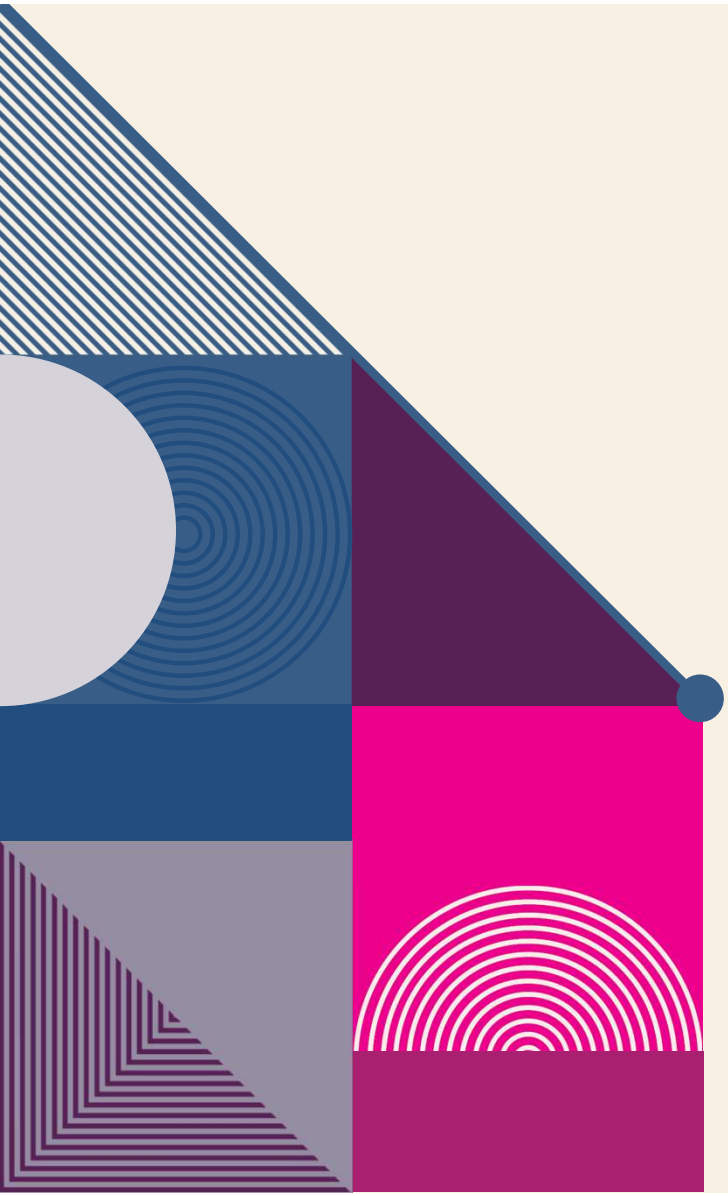
LEAD SCORING CASE STUDY

Soujanya Venugopal



PROBLEM STATEMENT

Problem Statement Summary: Lead Conversion Rate Improvement for X Education

X Education, an online course provider, faces a challenge in optimizing its lead conversion process. Despite generating numerous leads daily, the conversion rate stands at a suboptimal 30%. The company aims to enhance this by identifying and prioritizing potential leads, called 'Hot Leads,' to increase the lead conversion rate to approximately 80%.

PROBLEM SPECIFICS & SOLUTION

CURRENT SCENARIO:

•X Education attracts leads through various channels, including website visits, form submissions, and referrals.
•The lead conversion process involves subsequent communication from the sales team.
•The existing conversion rate is around 30%, indicating a significant room for improvement.

GOALS:

•Develop a logistic regression model to assign lead scores ranging from 0 to 100.
•A higher score indicates a higher likelihood of conversion, enabling the sales team to focus efforts on potential leads.

DATASET OVERVIEW:

•A dataset comprising 9000 data points is provided, containing attributes such as Lead Source, Total Time Spent on the Website, and Last Activity.
•The target variable, 'Converted,' denotes whether a lead resulted in a conversion (1) or not (0).

DATA CHALLENGES:

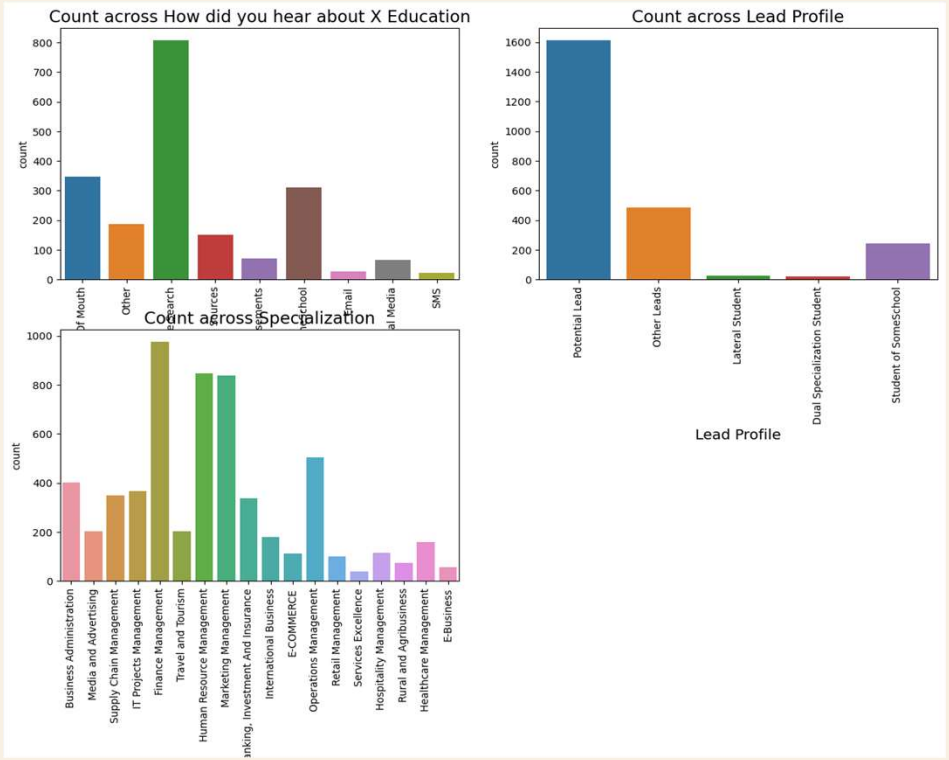
•Handle categorical variables with a level 'Select,' considering it as a null value.
•Explore and preprocess the dataset, addressing missing values and performing feature engineering.

PROBLEM APPROACH

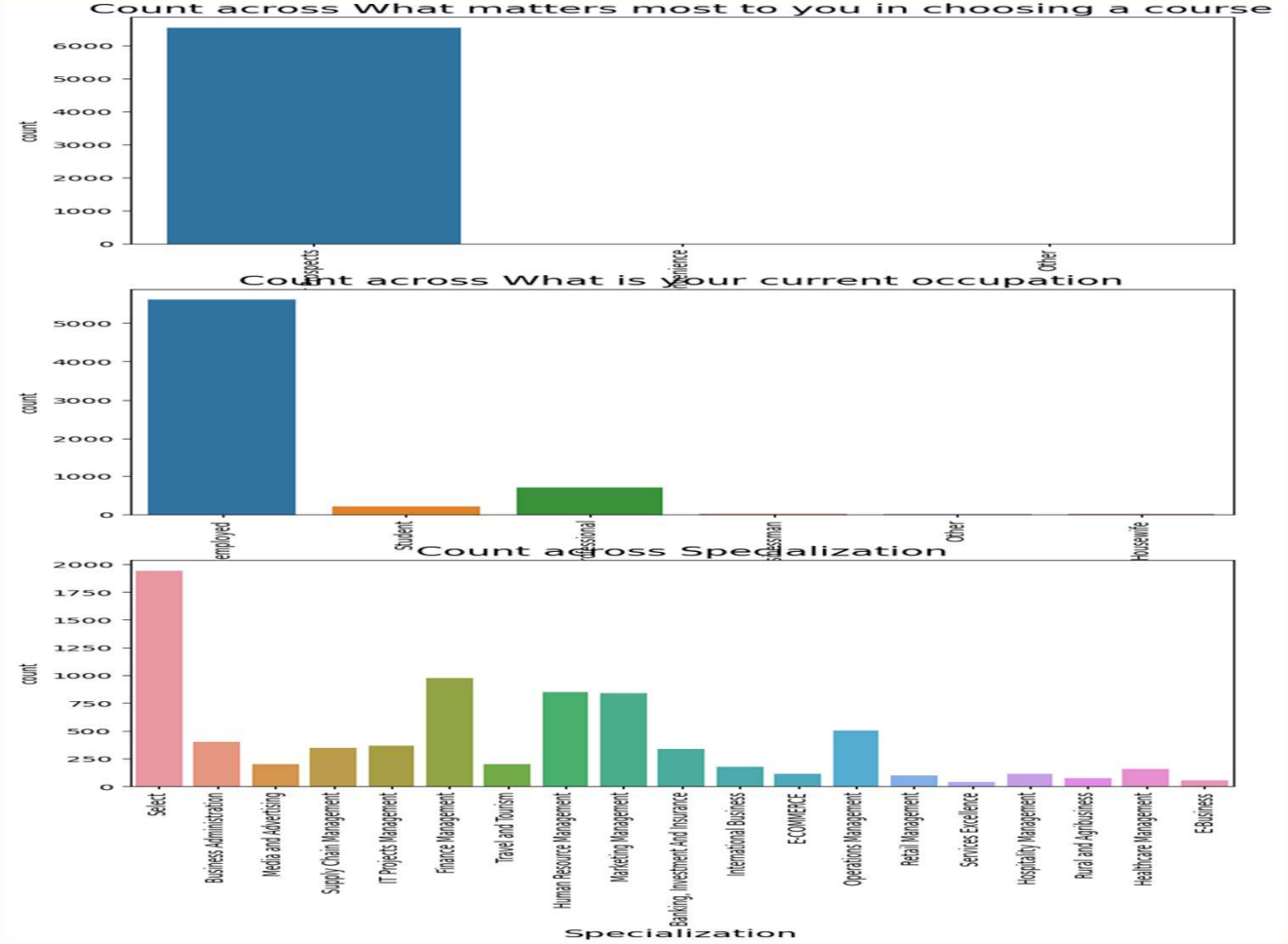
- Importing Data and Data Frame Inspection
- Preparing data
- Exploratory Data Analysis
- Dummy Variable Creation
- Test – Train Split
- Scaling Feature
- Correlation
- Model Building – RFE Rsquared VIF & p-values
- Model Evaluation
- Predicting Test Set

EDA – DATA CLEANING

Some columns in our data have a value called 'Select,' indicating that the student didn't choose any option for that particular category. This 'Select' is similar to having no information, like a blank answer. To understand how common this is, we want to count how many times 'Select' appears in each of these columns. This helps us identify how often students haven't chosen an option in those specific categories.

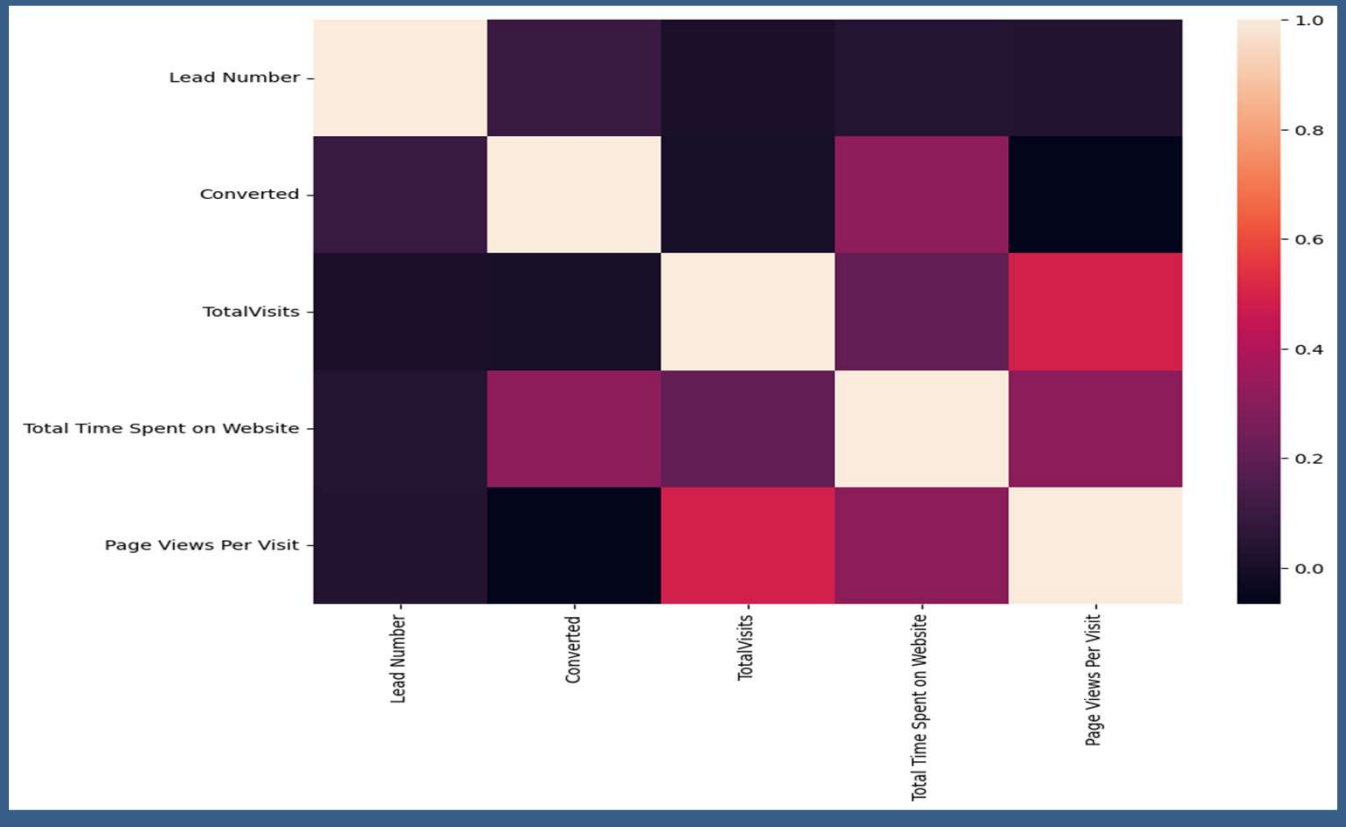


VISUALIZING THE FEATURES



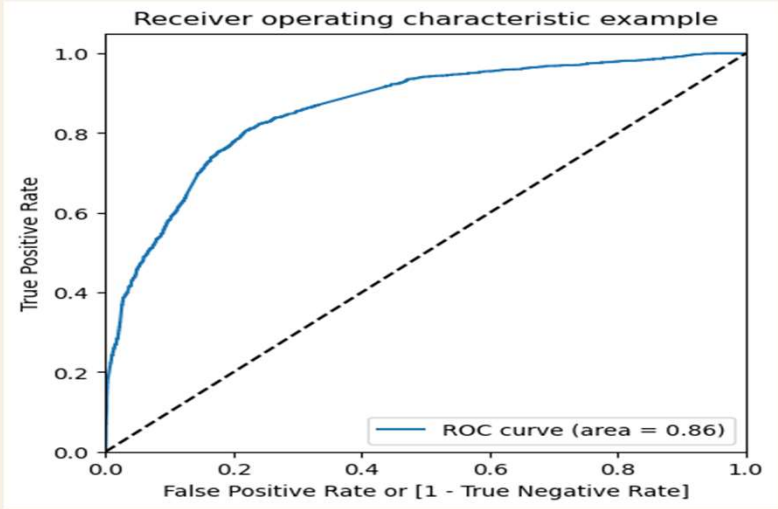
CORRELATION

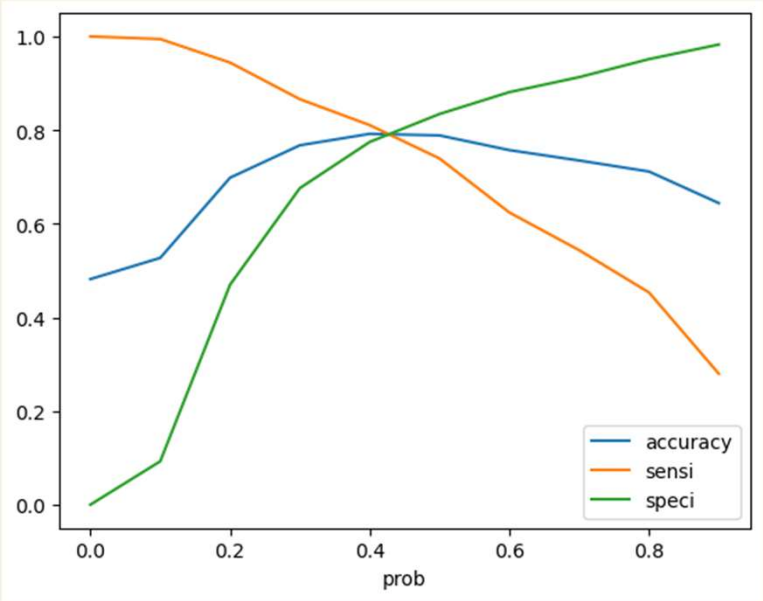
There is no Correlation between Variables

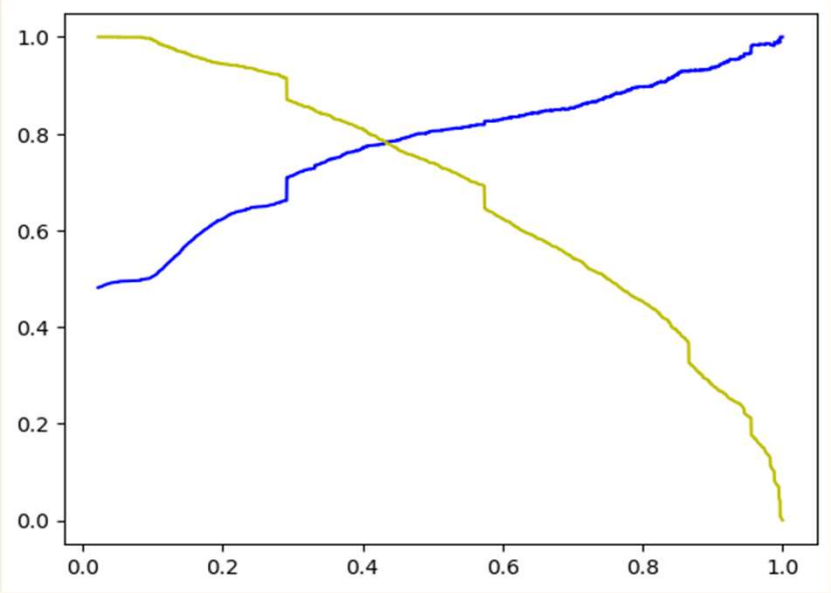


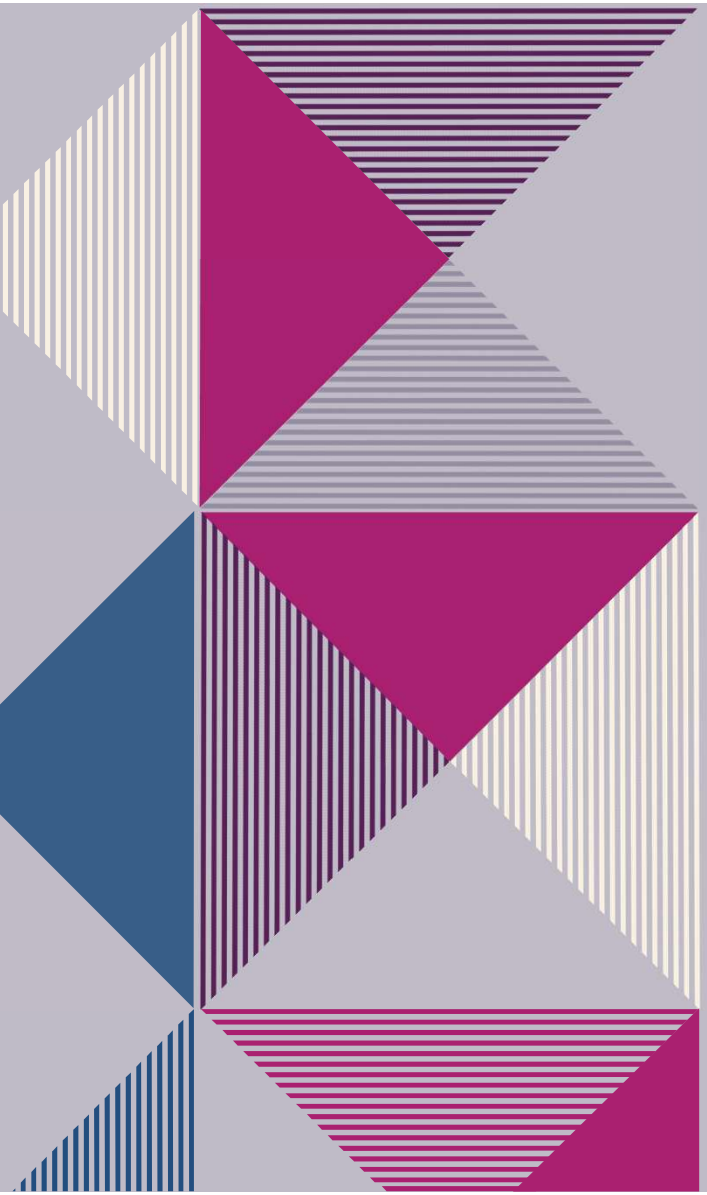

MODEL
EVALUATION

The ROC curve's Area Under the Curve (AUC) is 0.86, indicating a robust model. Now, let's examine the balance between sensitivity and specificity to pinpoint the best cutoff point. This tradeoff assessment will help us determine the optimal threshold, ensuring an effective balance between correctly identifying positives and negatives.









SUMMARY

In the early stages of lead generation, numerous potential customers are identified, but only a small portion eventually become paying clients. To improve lead conversion, it's crucial to nurture potential leads effectively during the middle stage. This involves educating them about the product, maintaining consistent communication, and sorting out the most promising prospects. Factors like 'TotalVisits,' 'Total Time Spent on Website,' and 'Page Views Per Visit' play a significant role in predicting a lead's conversion probability.

Once you've identified the best prospects, it's essential to maintain a list of these leads. This list serves as a valuable resource for informing them about new courses, services, job offers, and future educational opportunities. Monitoring each lead closely allows you to tailor the information you send based on their interests. Developing a thoughtful plan to address the unique needs of each lead contributes to successful lead capture.

Focusing on converted leads is key. Engage in question-and-answer sessions with leads to gather essential information. Through further inquiries and appointments, determine their intentions and readiness to join online courses. This proactive approach enhances the understanding of each lead, facilitating more effective communication and increasing the likelihood of successful conversions.


THANK YOU