Lead Scoring

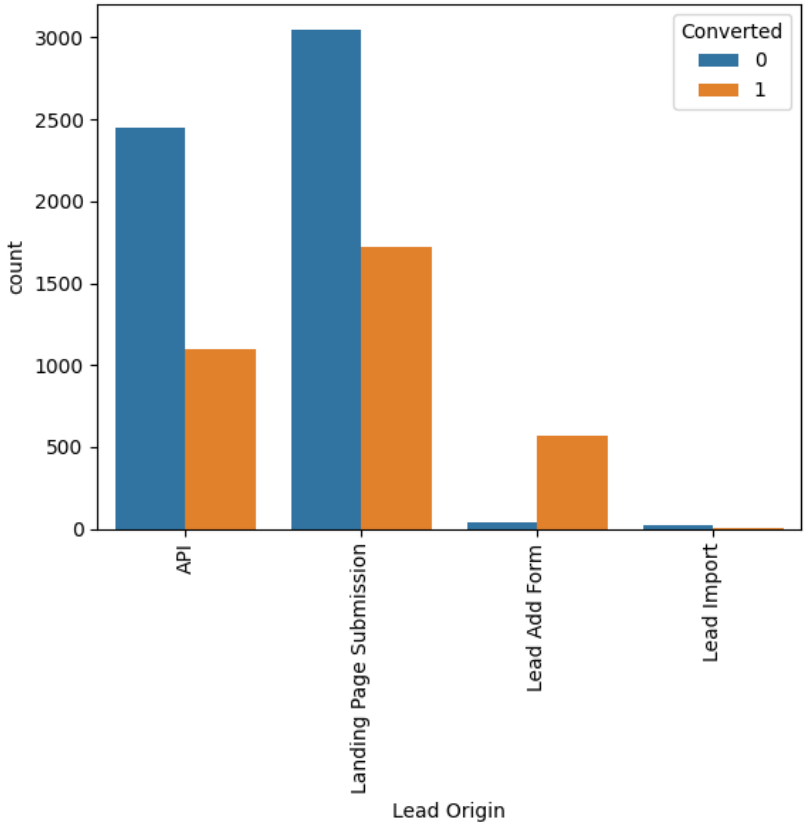
Problem Statement

The education company X Education is facing challenges in converting leads into paying customers. They receive a significant number of leads daily, but their lead conversion rate is low, currently at around 30%. To improve this conversion rate and identify the most potential leads, X Education has appointed us to build a lead scoring model.

The objective is to assign a lead score to each lead, indicating their likelihood of converting into a paying customer. By focusing more on leads with higher scores, the company aims to increase its lead conversion rate to approximately 80%.

Analysis

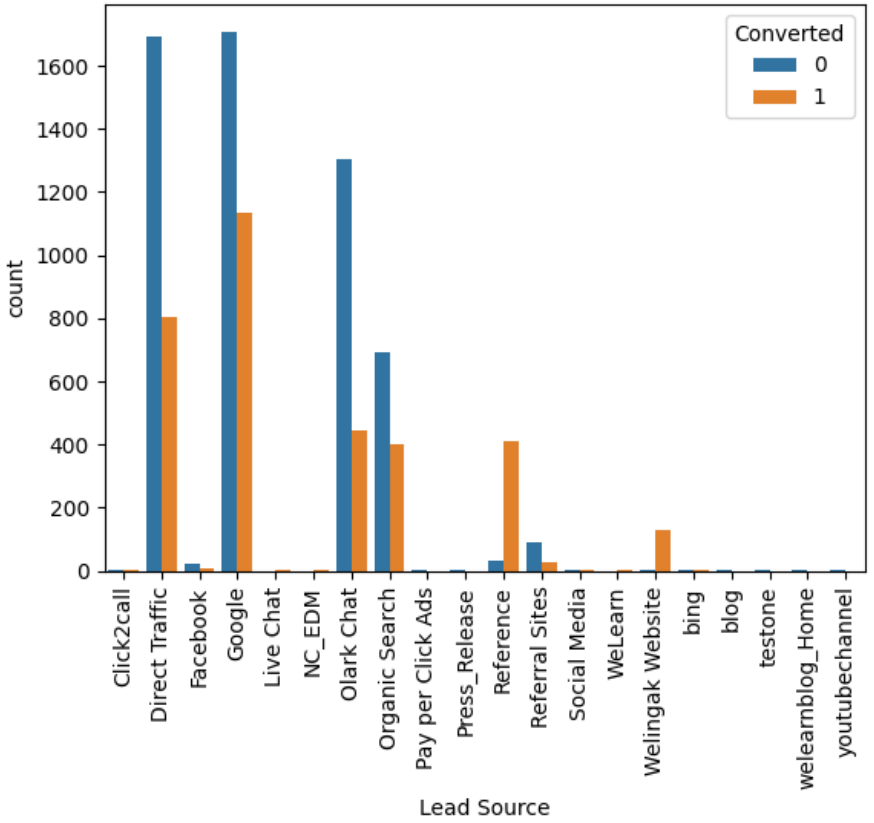
Analyse ‘Lead Origin’



- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import are very less in count.

=> To improve overall lead conversion rate, we need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

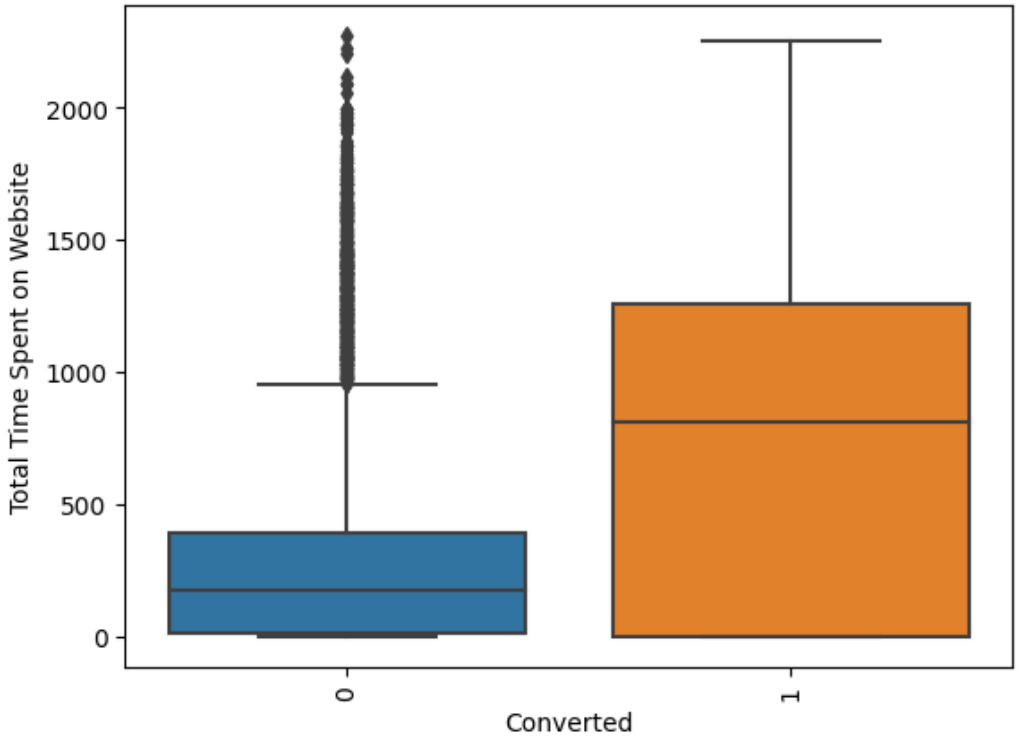
Analyse ‘Lead Source’



- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.

=> To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.

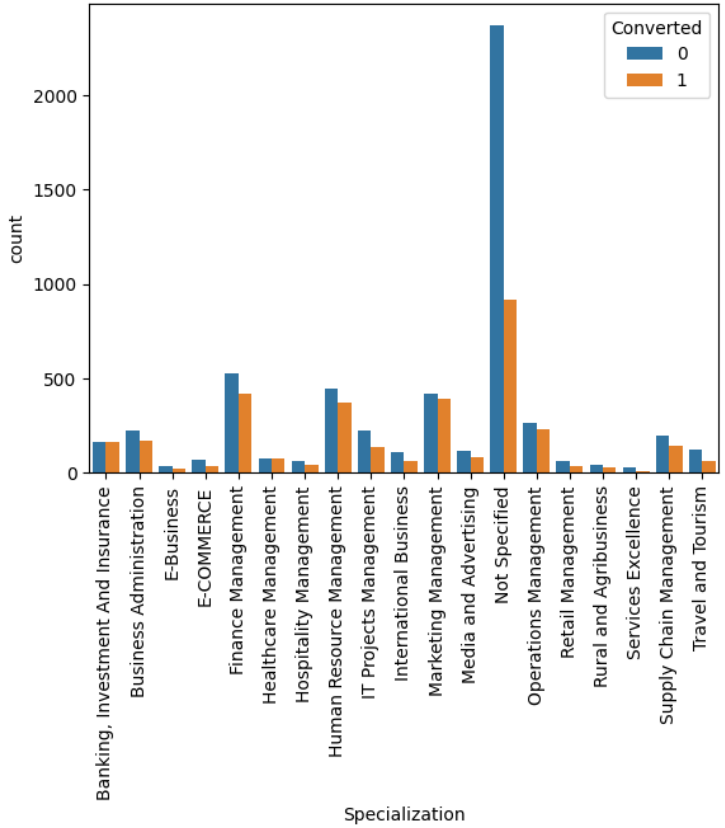
Analyse ‘Total Time Spent on Website’



- Leads spending more time on the website are more likely to be converted.

  => Website should be made more engaging to make leads spend more time.

Analysing ‘Specialization’



=> Focus should be more on the Specialization with high conversion rate.

Methodology

Methodology

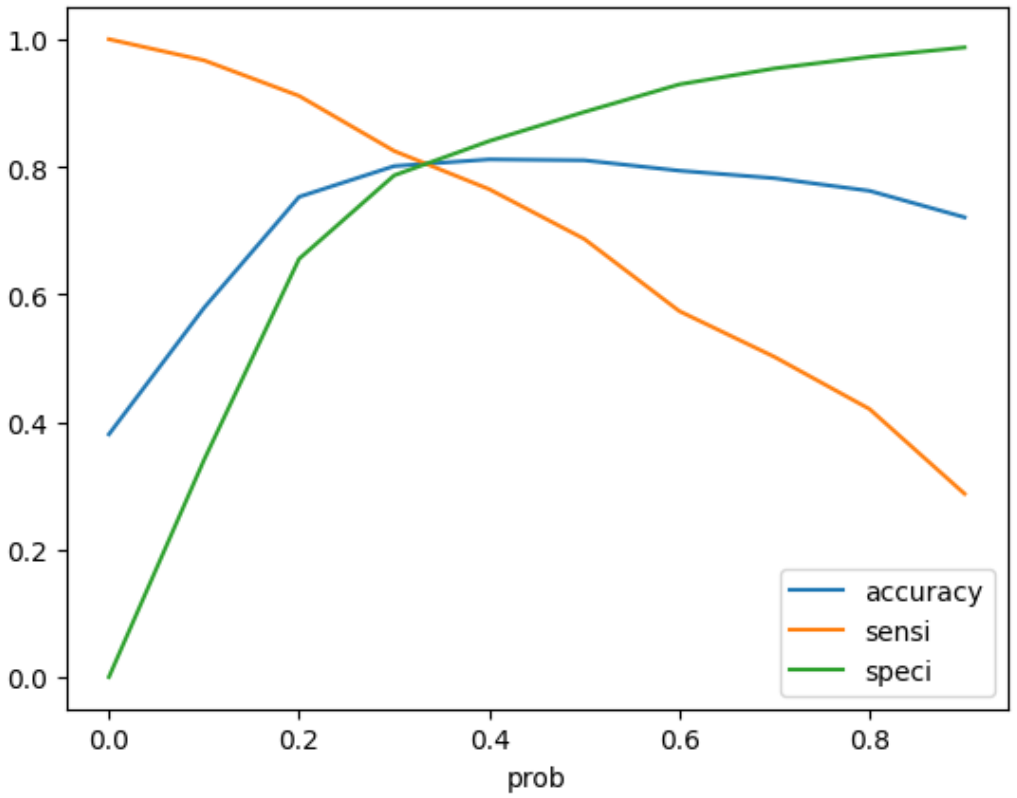
The model development process involves several steps:

1. Data Understanding: We will thoroughly analyze the provided dataset, which includes information about leads, their interactions, and conversion outcomes.
2. Data Preprocessing: Cleaning the dataset, handling missing values, and treating categorical variables will be essential to ensure accurate and consistent data for analysis.
3. Feature Engineering: We will enhance the predictive power of the model by creating relevant features, encoding categorical variables, and deriving meaningful insights from the available data.
4. Model Training: Logistic regression, a popular classification algorithm, will be employed to build the lead scoring model. This algorithm is well-suited for binary classification problems and provides interpretability.
5. Model Evaluation: The trained model will be evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. This evaluation will measure the model's performance in predicting lead conversions.
6. Lead Scoring and Conversion Prediction: Once the model is trained and validated, it will be applied to assign lead scores to new leads. The scores will help prioritize efforts towards potential customers who are more likely to convert.

Selecting features

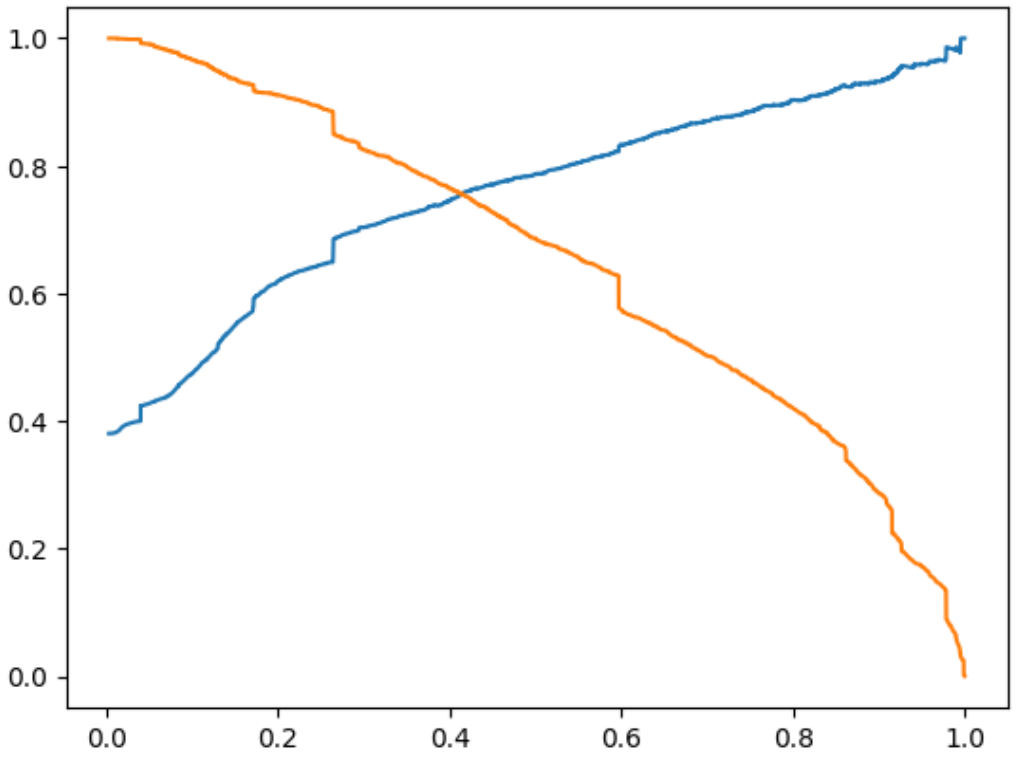
- We should make calls to the leads coming from:
  - Lead Origin: Lead Add Form
  - What is your current occupation: Working Professional
  - Longer total time spent on Website
  - Last Activity: SMS Sent
- We should not make calls to the leads coming from:
  - Last Notable Activity: Email Opened, Page Visited on Website, Modified, Email Link Clicked
  - Lead Source: Google, Organic Search, Referral Sites, Direct Traffic
  - Do Not Email: yes
  - Last Activity: Olark Chat Conversation

Sentitivity and Specificity



- Accuracy: 81%
- Sensitivity: 80 %
- Specificity: 81 %
- False Positive rate: 19%
- Positive Predictive Value: 72%
- Negative Predictive Value: 87%

Precision and Recall



- Precision : 79%
- Recall : 69%

Result

- Train data:
  - Accuracy: 81%
  - Sensitivity: 80%
  - Specificity: 81%
- Test data:
  - Accuracy: 80%
  - Sensitivity: 80%
  - Specificity: 80%

- The model can predict the Conversion Rate well to help making calls with a higher lead conversion rate of 80%
- There are 321 hot leads with score >= 90