LEAD SCORING
GROUP CASE STUDY

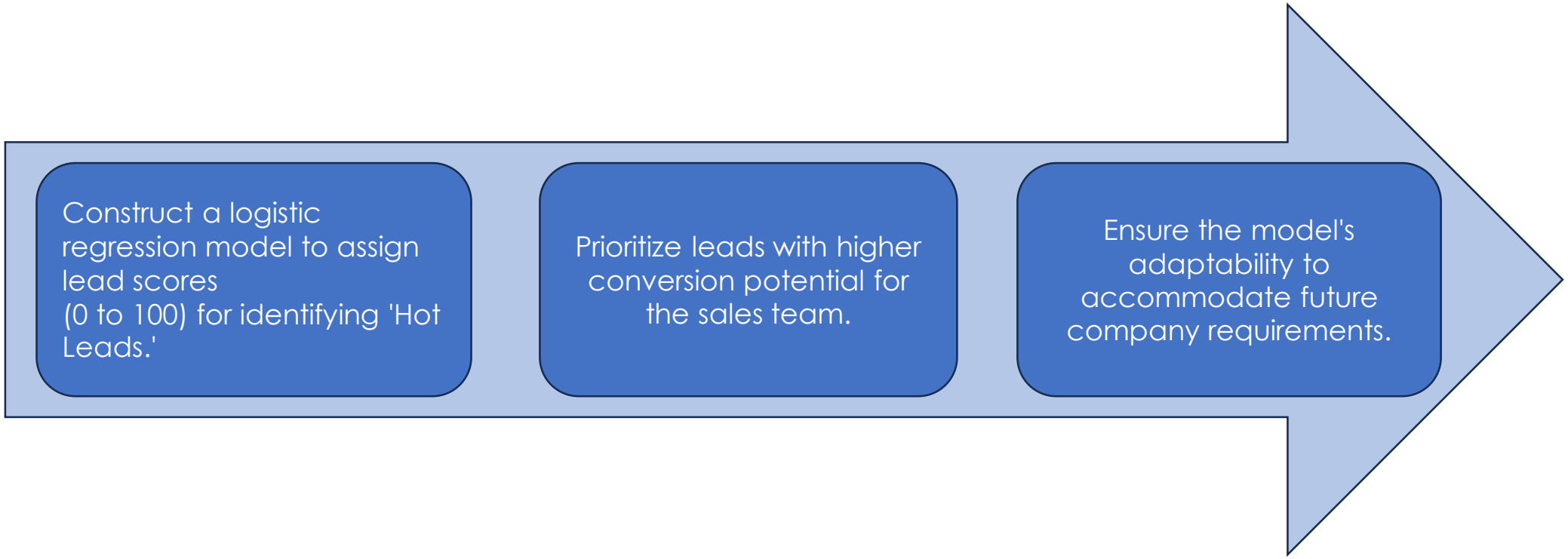
ARJIT AREN
ARIJIT PAUL
ARINDAM DAS
BATCH (DSC59 August 2023)

PROBLEM STATEMENT

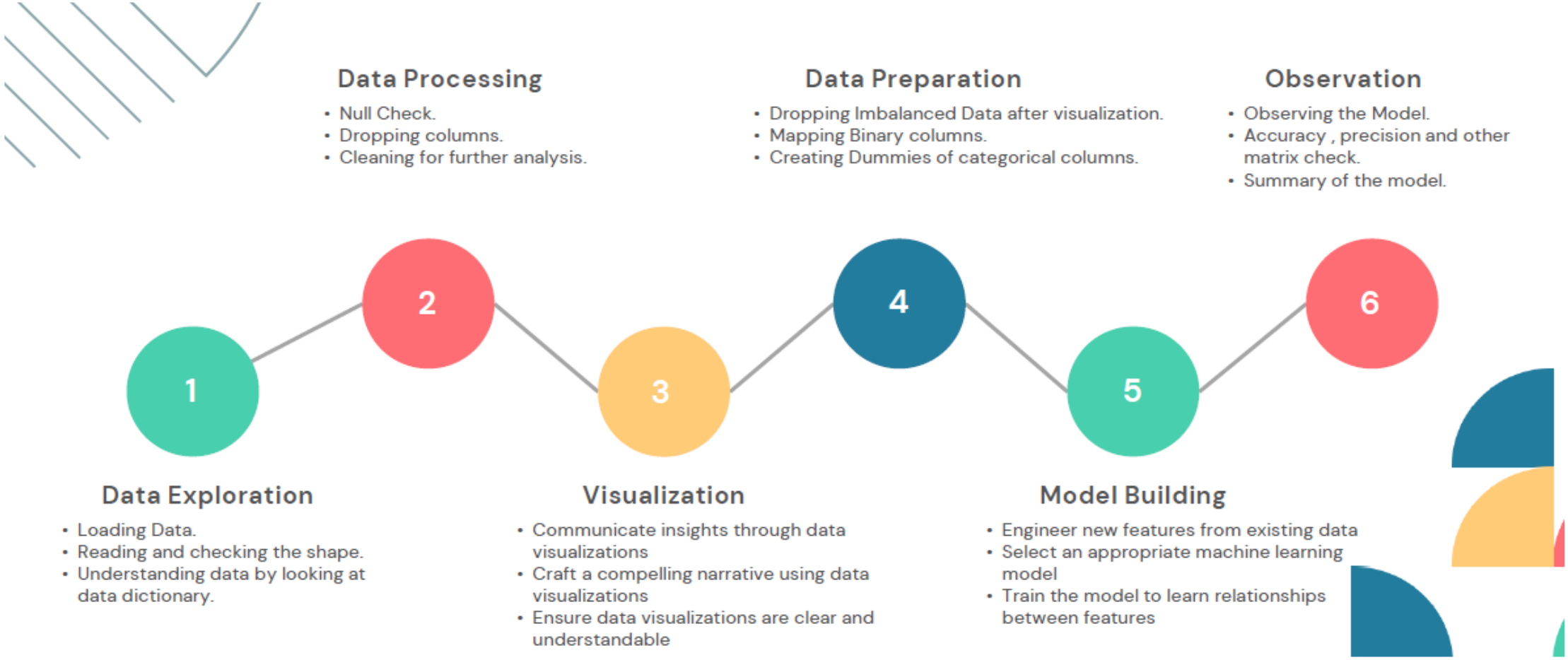
- The online course provider X Education has a problem with its lead conversion rate even though it is obtaining a significant amount of leads. Thirty percent or so of generated leads end up becoming paying clients. The organization prioritizes and identifies "Hot Leads," or those with a better chance of converting, in an effort to increase this conversion rate.

- Building a logistic regression model with lead scores ranging from 0 to 100 is the goal. With this model, the business may concentrate its sales efforts on leads who have a higher chance of converting while still guaranteeing flexibility in response to future demand changes. Included in the dataset are variables like Lead Source, Total Time Spent on Website, and Last Activity, among others. The target variable 'Converted' represents previous conversions.

OBJECTIVE


Construct a logistic regression model to assign lead scores (0 to 100) for identifying 'Hot Leads.'
Prioritize leads with higher conversion potential for the sales team.
Ensure the model's adaptability to accommodate future company requirements.

APPROACH


Data Processing
• Null Check.
• Dropping columns.
• Cleaning for further analysis.
Data Preparation
• Dropping Imbalanced Data after visualization.
• Mapping Binary columns.
• Creating Dummies of categorical columns.
Observation
• Observing the Model.
• Accuracy , precision and other matrix check.
• Summary of the model.
1
2
3
4
5
6
Data Exploration
• Loading Data.
• Reading and checking the shape.
• Understanding data by looking at data dictionary.
Visualization
• Communicate insights through data visualizations
• Craft a compelling narrative using data visualizations
• Ensure data visualizations are clear and understandable
Model Building
• Engineer new features from existing data
• Select an appropriate machine learning model
• Train the model to learn relationships between features

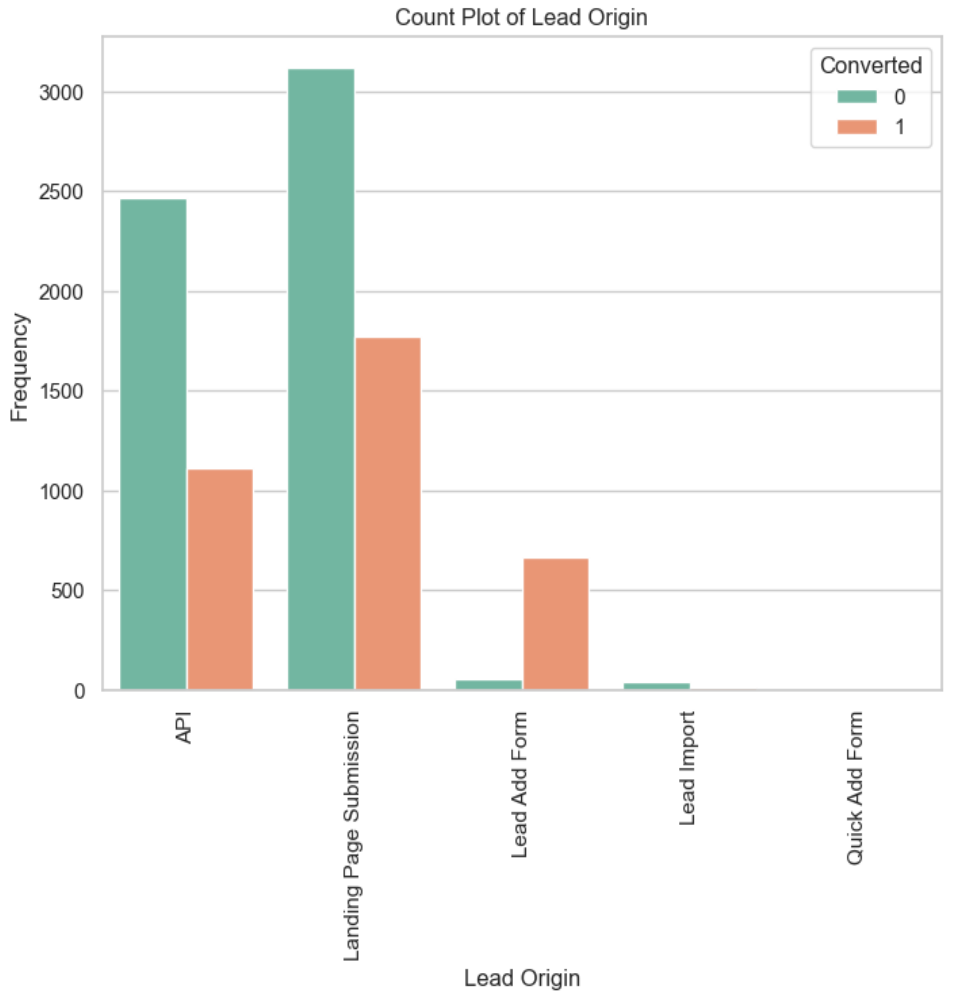
Lead Origin

1. Most leads are customers that were found through submissions to Landing Pages.

2. Lead Add Form-originated clients are fewer in quantity, but they have a higher conversion rate.

3. Lead Import and Lead Origin from API & Lead Import have the lowest conversion rates, with Lead Import having few clients.

In order to increase the overall conversion rate, concentrate on increasing lead generation through Lead Add Forms and optimizing conversions from API and Landing Page submissions.

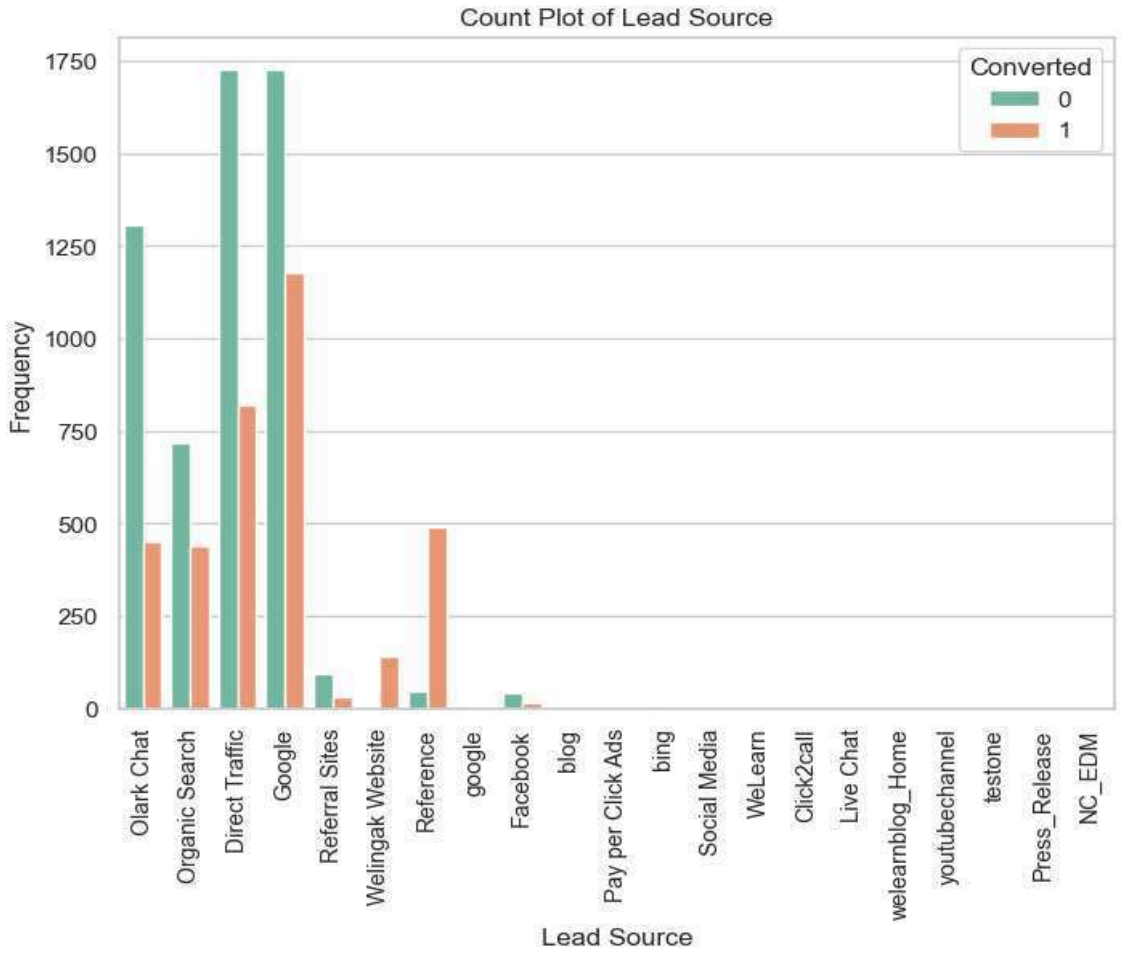


Lead Source

1. The two main lead sources are direct traffic from Google.

2. Leads obtained through Google have the highest conversion rate.

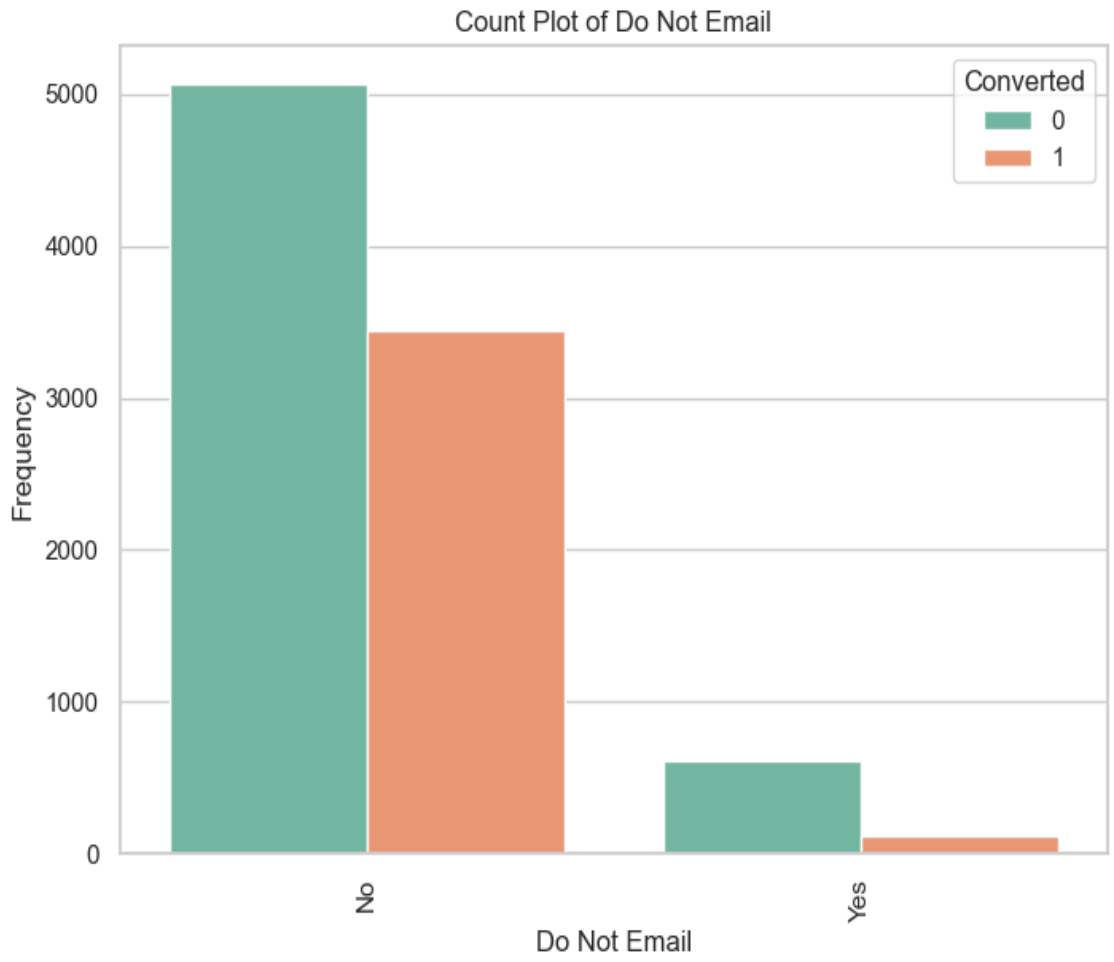
3. There is a significant chance that leads from the WelingkarWebsite and Reference will convert.



Do Not Email

1. The conversion rate of customers who choose to "Do Not Mail" is lower.

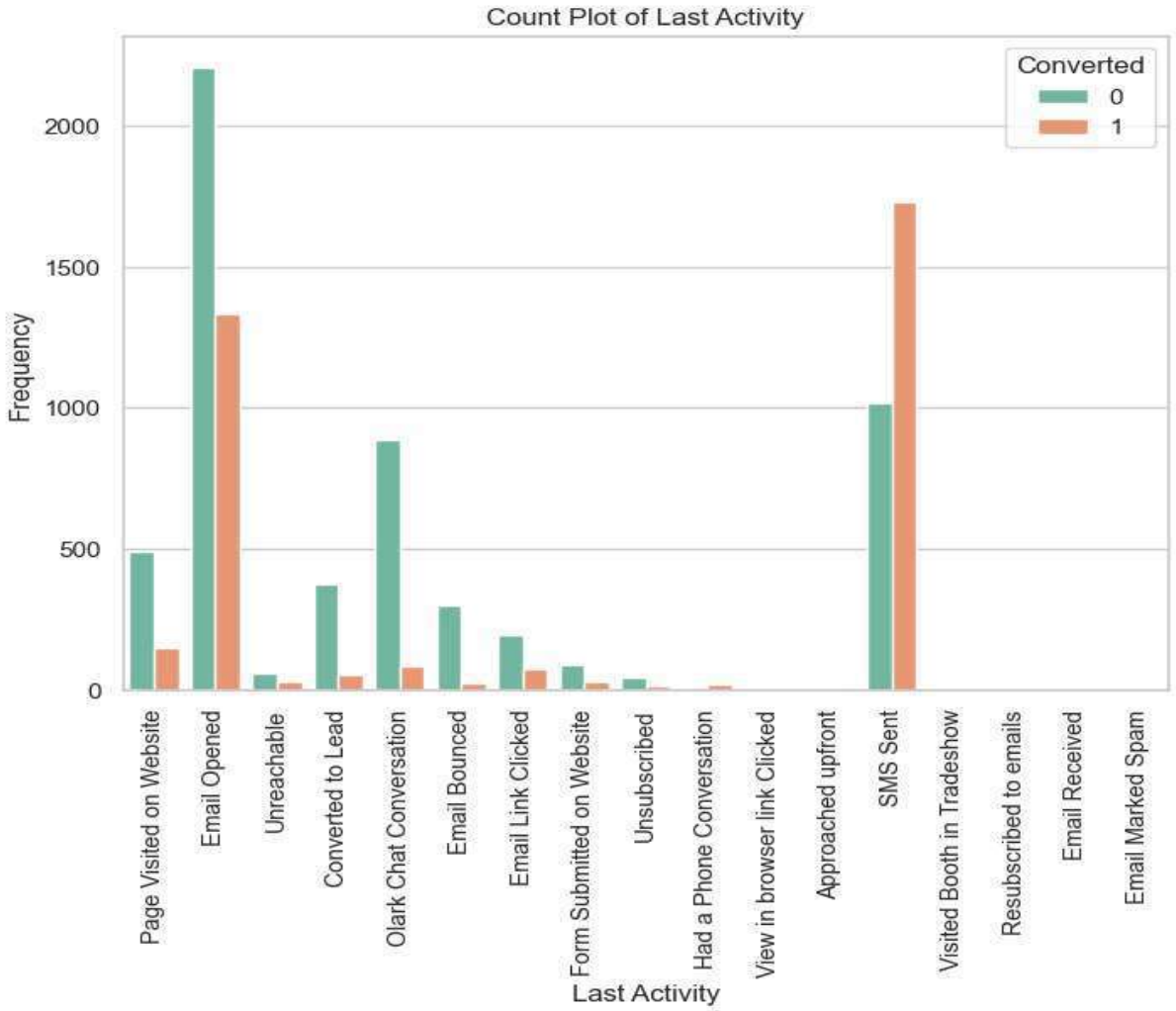
2. The majority of leads come from consumers with a better conversion rate who do not choose to do not mail.



Last Activity

1. The conversion rate is higher for customers who sent an SMS as their most recent activity.

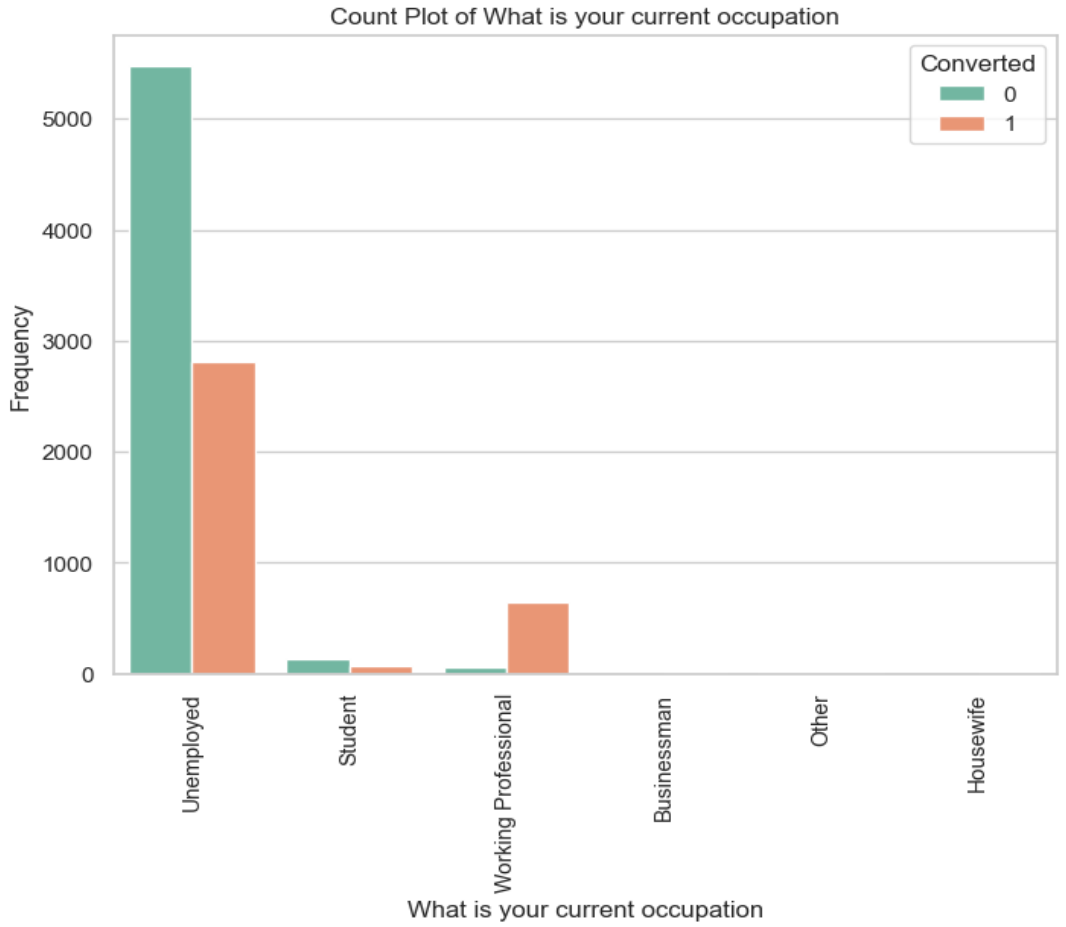
2. The majority of customers, with a conversion rate, have the most recent activity as Email Opened.Increase the number of leads from consumers whose last activity was SMS Sent and concentrate on increasing conversions from those whose last activity was email opened in order to raise the overall conversion rate.



Current Occupation

1.The majority of leads are jobless.

2. Professionals in the workforce are those with conversion rates exceeding 90%.

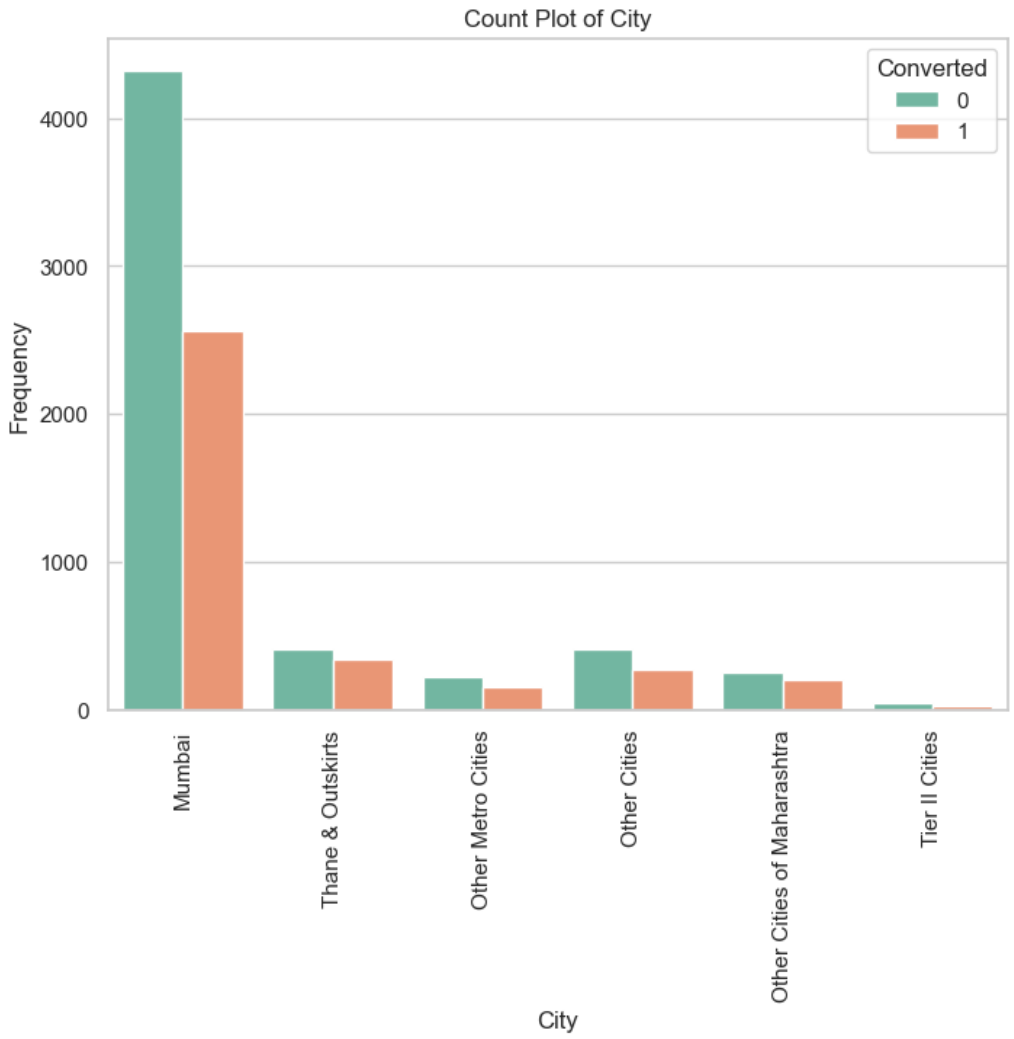


City

1.Thane & Outskirts have the highest conversion rate, but Mumbai generates the most leads.

2.Other metro cities have a higher conversion rate than the total, which suggests that concentrating your marketing efforts in these areas would be a good idea.

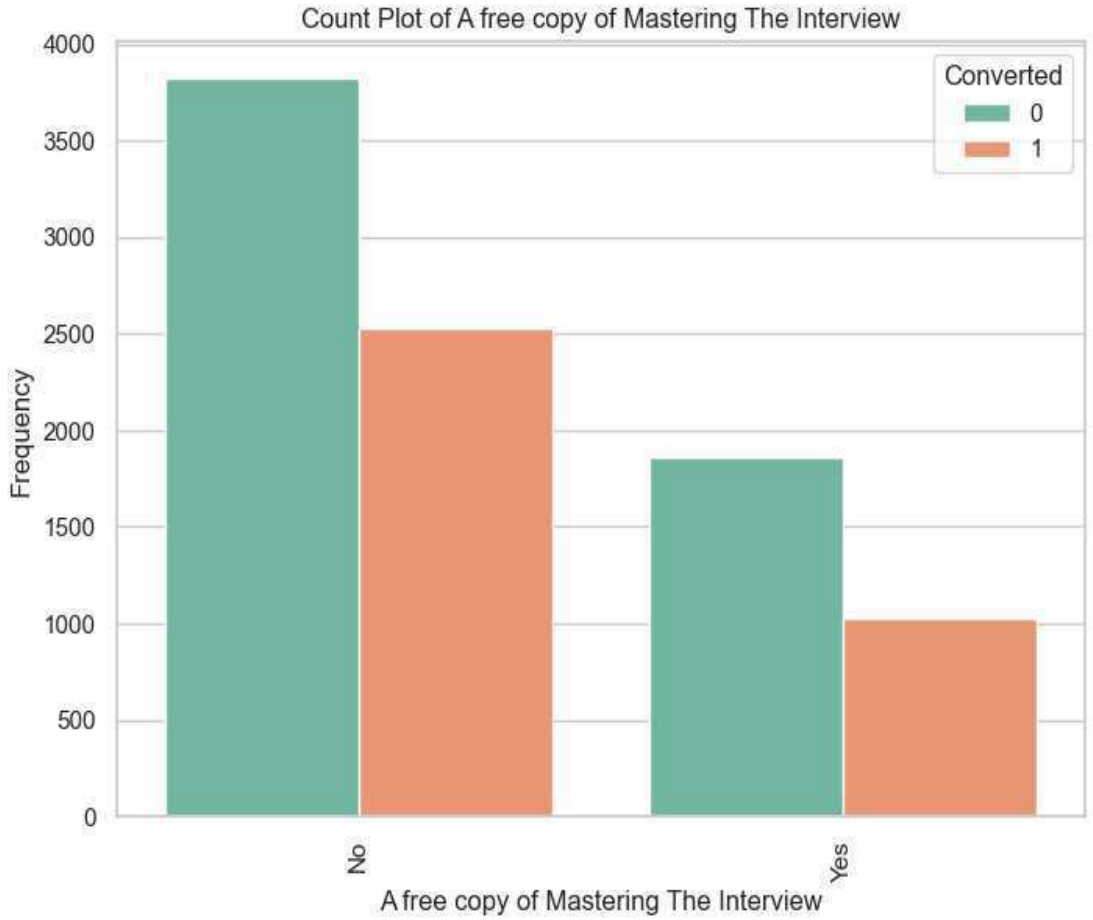
3. Compared to the national conversion rate, the rates for Other Cities and Other Cities in Maharashtra are lower. 4.Tier II Cities have the lowest conversion rates, indicating that it would be harder to convert leads in this market.



A free copy of Mastering The Interview

1.A free copy of Mastering the Interview was offered to leads, and leads who took it were more likely to convert than those who did not. Leads who received a complimentary copy of the book had a conversion rate of 35%, whereas leads who did not had a conversion rate of only 20%.

2.This implies that giving away a free copy of Mastering the Interview to leads can be a highly successful strategy for raising conversion rates.

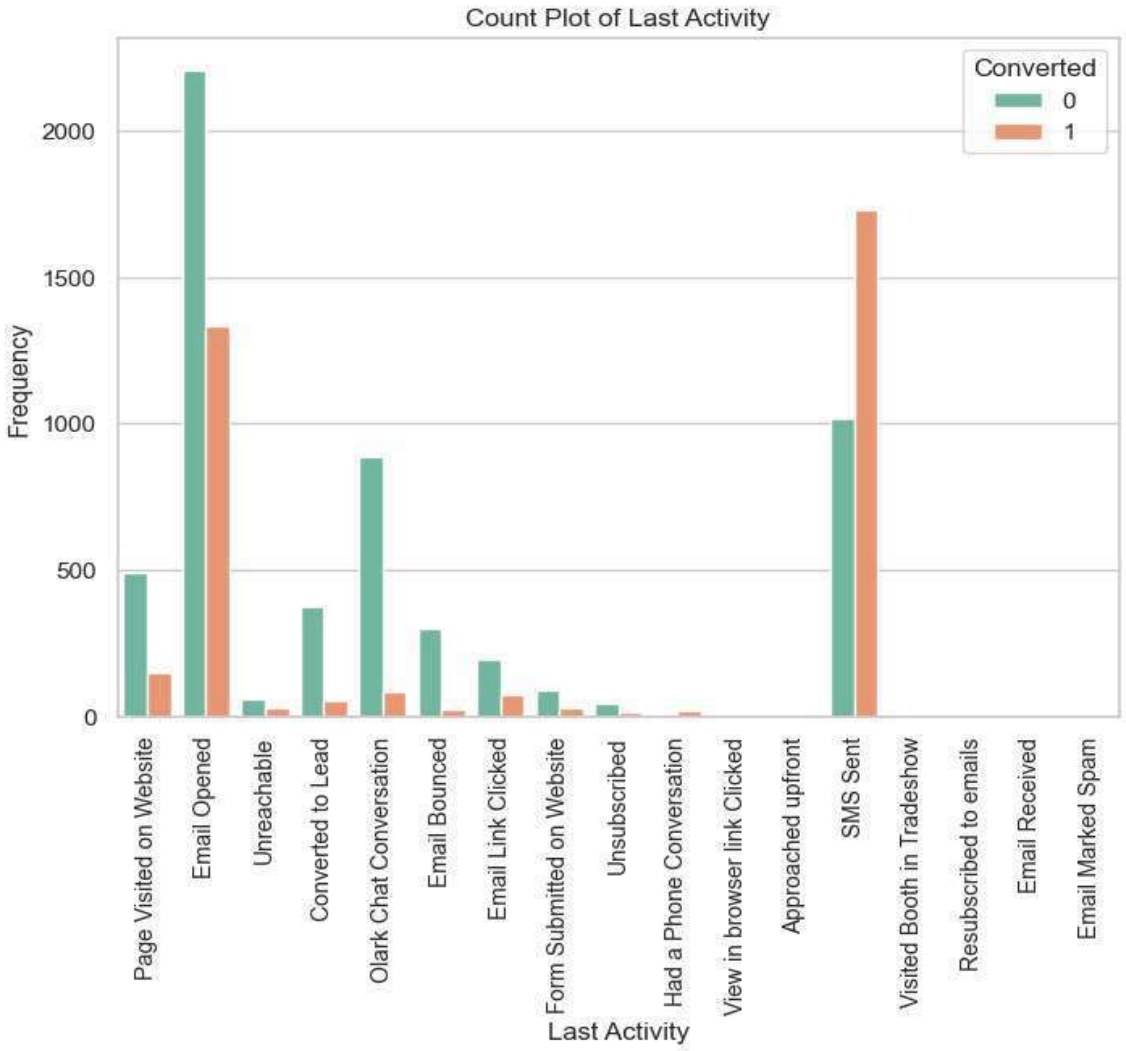


Last Notable Activity

1.Numerous leads that were converted have "Modified" their contact details. This implies that since converted leads are more likely to modify their contact information than non-converting leads, it is critical to maintain an up-to-date lead database.

2.The email addresses of a sizable portion of converted leads are marked as "Unreachable" or "Unsubscribed". This implies that it's critical to regularly purge incorrect email addresses and unsubscribers from your email list.

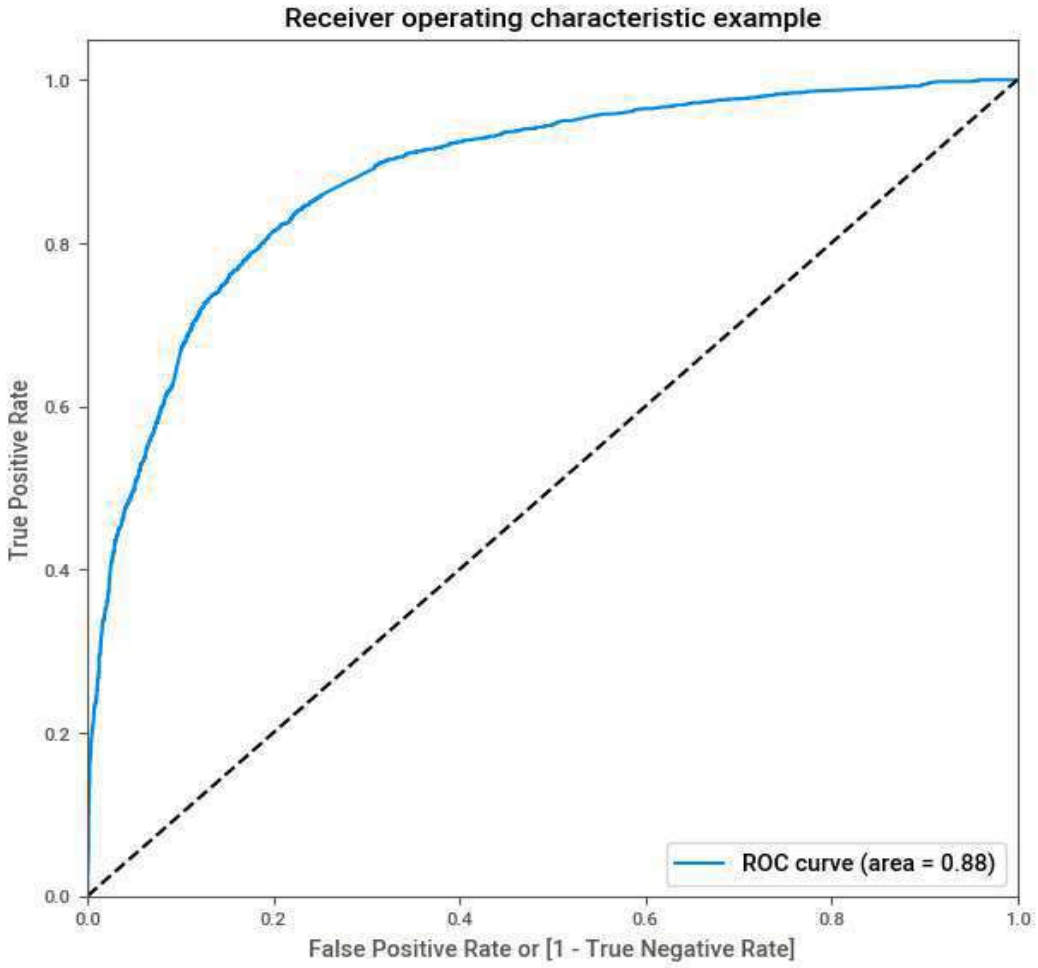
3."Had a Phone Conversation" and "Olark Chat Conversation" are present in a comparatively small percentage of converted leads. This implies that alternative channels, such website visits or email, might be more effective at converting leads than these ones.



ROC Curve

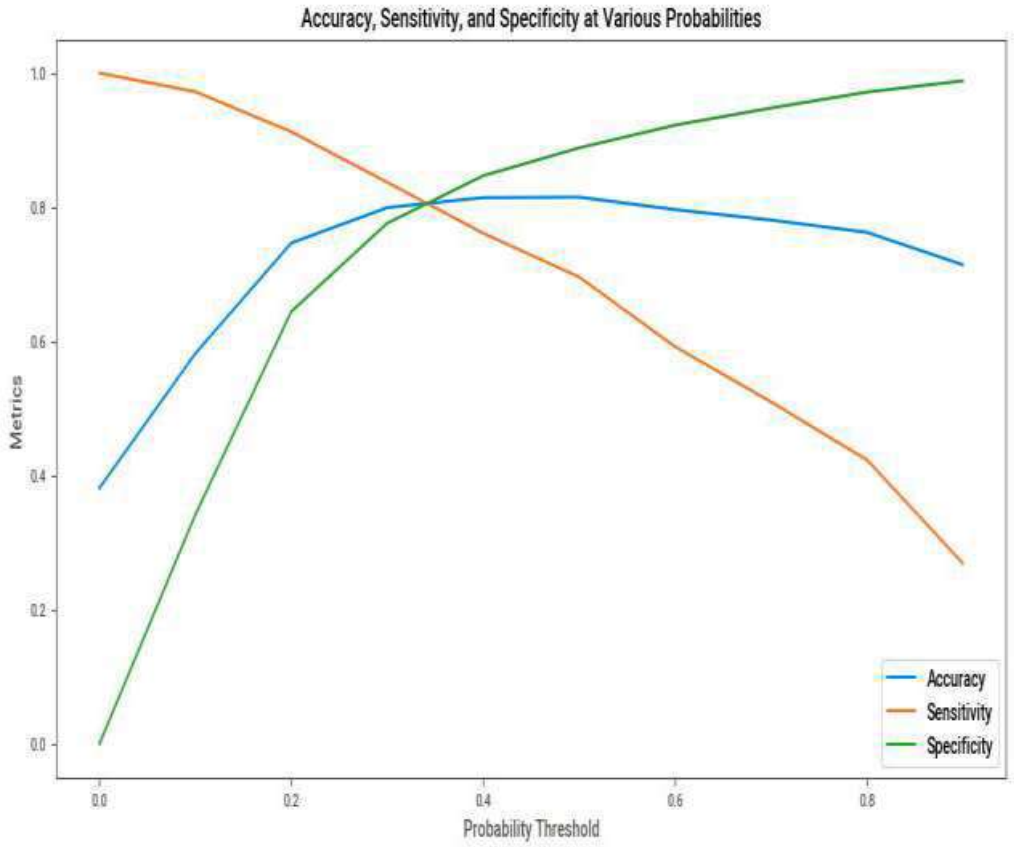
1.The area under the curve (AUC) value of 0.88 indicates that the ROC curve in this instance has a good degree of accuracy.

2.This suggests that the classifier can reduce the amount of false positives while accurately recognizing most positive cases.



Accuracy, Sensitivity, and Specificity at Various Probabilities

- 1. The relationship between specificity, sensitivity, and accuracy at different probabilities is displayed on the graph.
- 2. The percentage of all forecasts that come true, whether they be positive or negative, is known as accuracy.
- 3. The percentage of affirmative cases that are accurately detected is known as sensitivity.
- 4. The percentage of negative cases that are accurately identified is known as specificity.



Model Evaluation

1. The model's improved performance on test data over training data indicates that it is adapting effectively to new data.

2. The model's precision is 81%, meaning that 81% of the time it correctly predicts a positive case.

3. The recall of the model is 81%, indicating that 81% of all true positive cases are identified by it.

4. The harmonic mean of precision and recall gives the model an F1-score of 81%. This implies that the model's precision and recall are well-balanced. In general, the model has strong performance on both training and test sets. It can recognize positive and negative instances with accuracy.

Training Data:

	Precision	Recall	F1-Score	Support
Class 0	0.86	0.83	0.85	4002
Class 1	0.74	0.77	0.76	2466
Accuracy			0.81	6468
Macro Avg	0.80	0.80	0.80	6468
Weighted Avg	0.81	0.81	0.81	6468

Test Data:

	Precision	Recall	F1-Score	Support
Class 0	0.85	0.85	0.85	1677
Class 1	0.77	0.76	0.77	1095
Accuracy			0.82	2772
Macro Avg	0.81	0.81	0.81	2772
Weighted Avg	0.82	0.82	0.82	2772

Conclusion

- Perspectives from the Data on Lead Conversion
- In conclusion, the information offered in this talk sheds light on lead behavior and conversion rate influencing variables including city, tag, and noteworthy activity. Conversion rates can be raised by creating focused interventions with the help of the insights gathered from this data.


Thank
You