LEAD SCORING CASE STUDY

Group Study

PROBLEM STATEMENT

X Education, an online course provider, faces a challenge with its lead conversion rate despite acquiring a substantial volume of leads. Approximately 30% of acquired leads get converted into paying customers. The company seeks to enhance this conversion rate by identifying and prioritizing 'Hot Leads,' those with a higher likelihood of conversion.

The aim is to construct a logistic regression model that assigns lead scores from 0 to 100. This model will enable the company to focus its sales efforts on leads more likely to convert while also ensuring adaptability to future changes in requirements. The dataset provided includes attributes such as Lead Source, Total Time Spent on Website, and Last Activity, among others, with the target variable 'Converted' indicating past 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

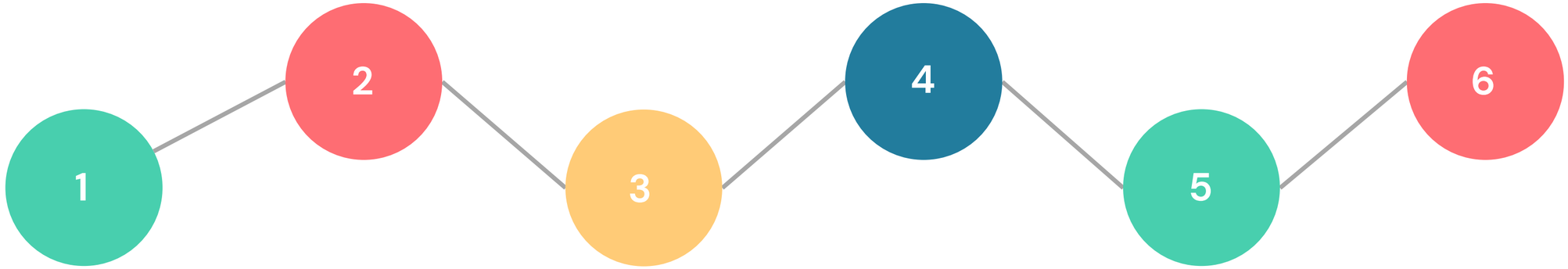
- Null Check.
- Dropping columns.
- Cleaning for further analysis.

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.



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

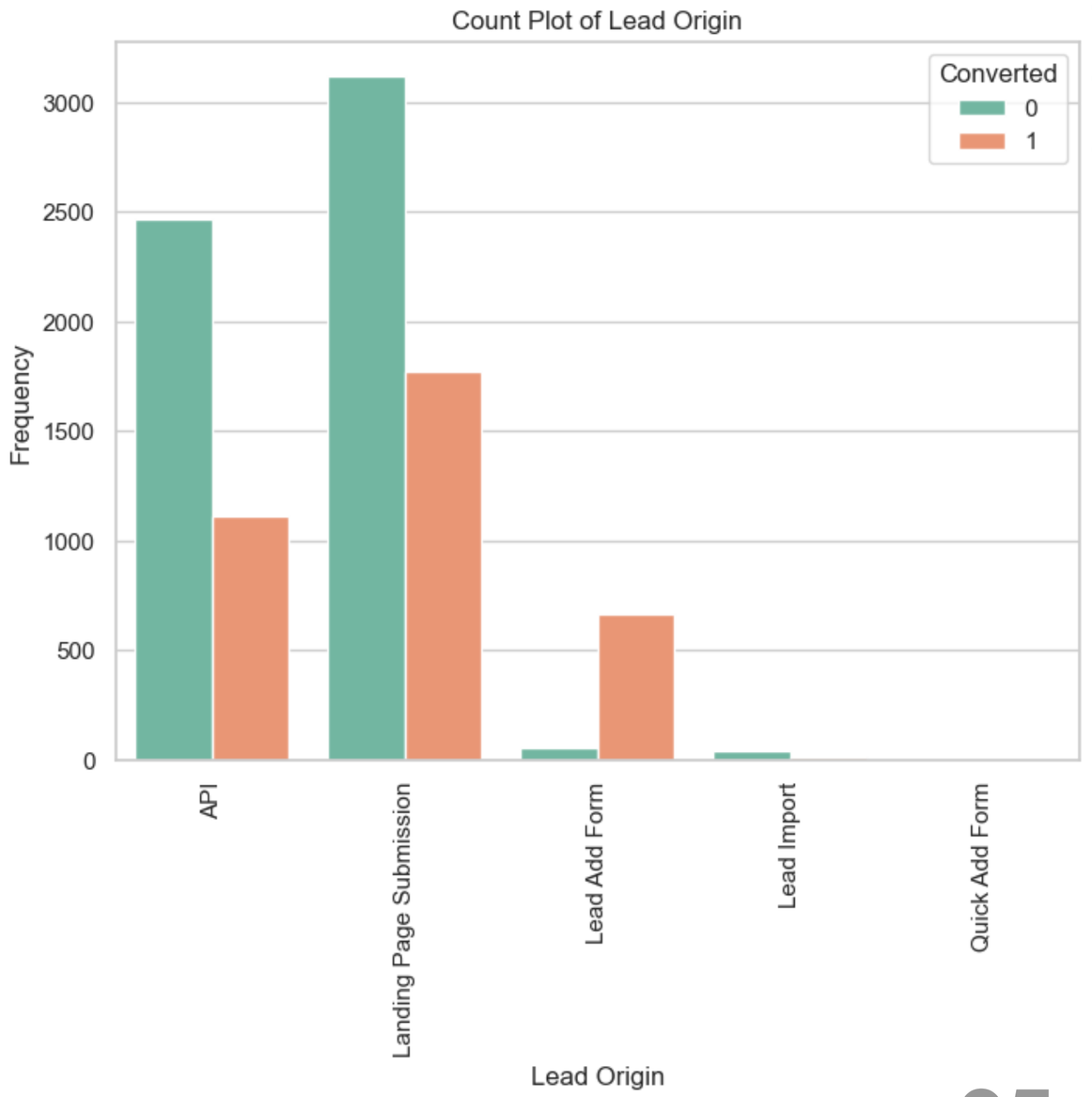
Lead Origin

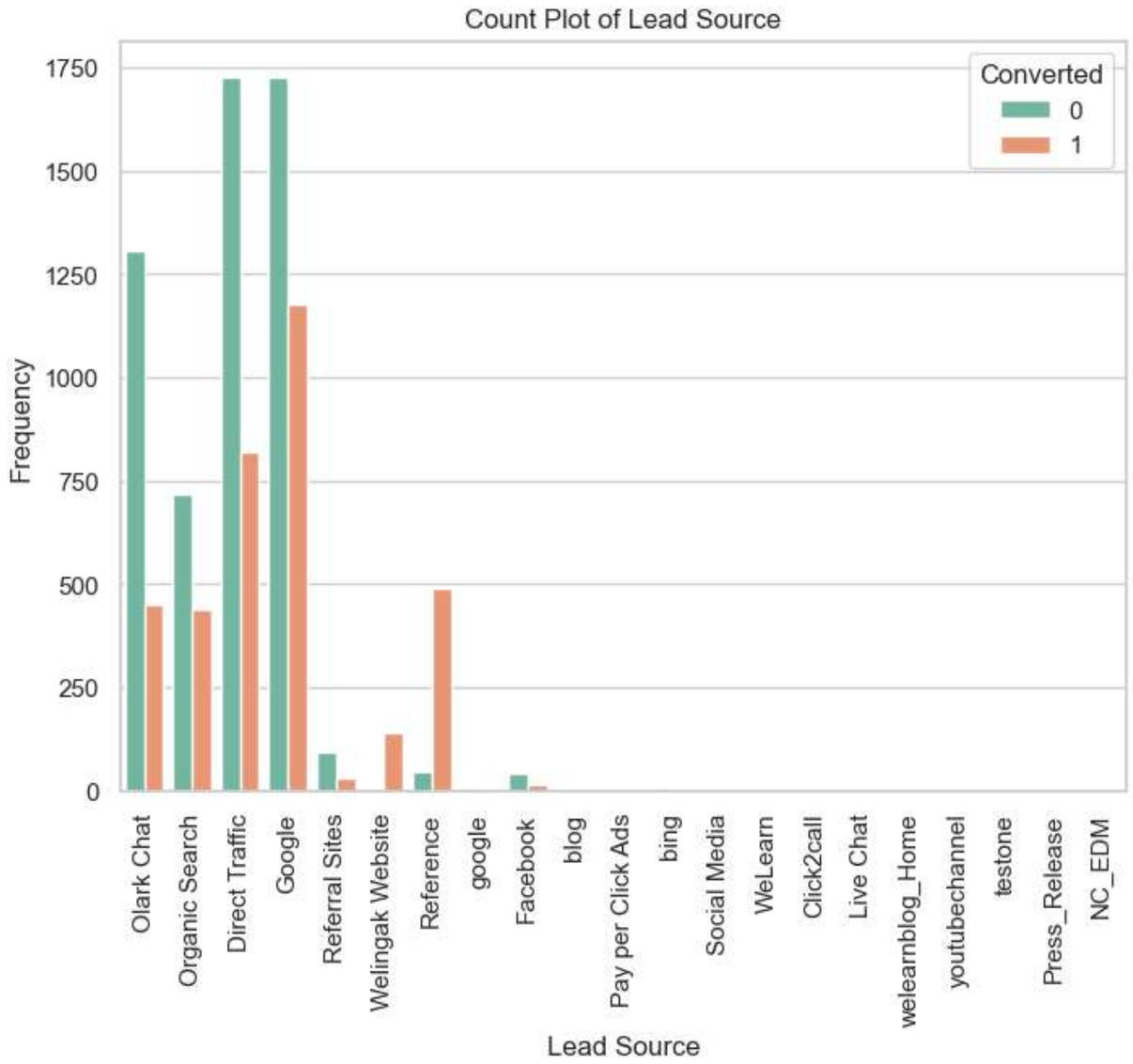
1. Customers identified from Landing Page submissions constitute the majority of leads.

2. Lead Add Form-originated customers show a high probability of conversion but are fewer in number.

3. Lead Origin from API & Lead Import has the least conversion rate, with few customers from Lead Import.

To enhance the overall conversion rate, focus on improving conversions from API and Landing Page submissions while generating more leads through Lead Add Forms.





Lead Source

1. Google & Direct Traffic are the major lead sources.

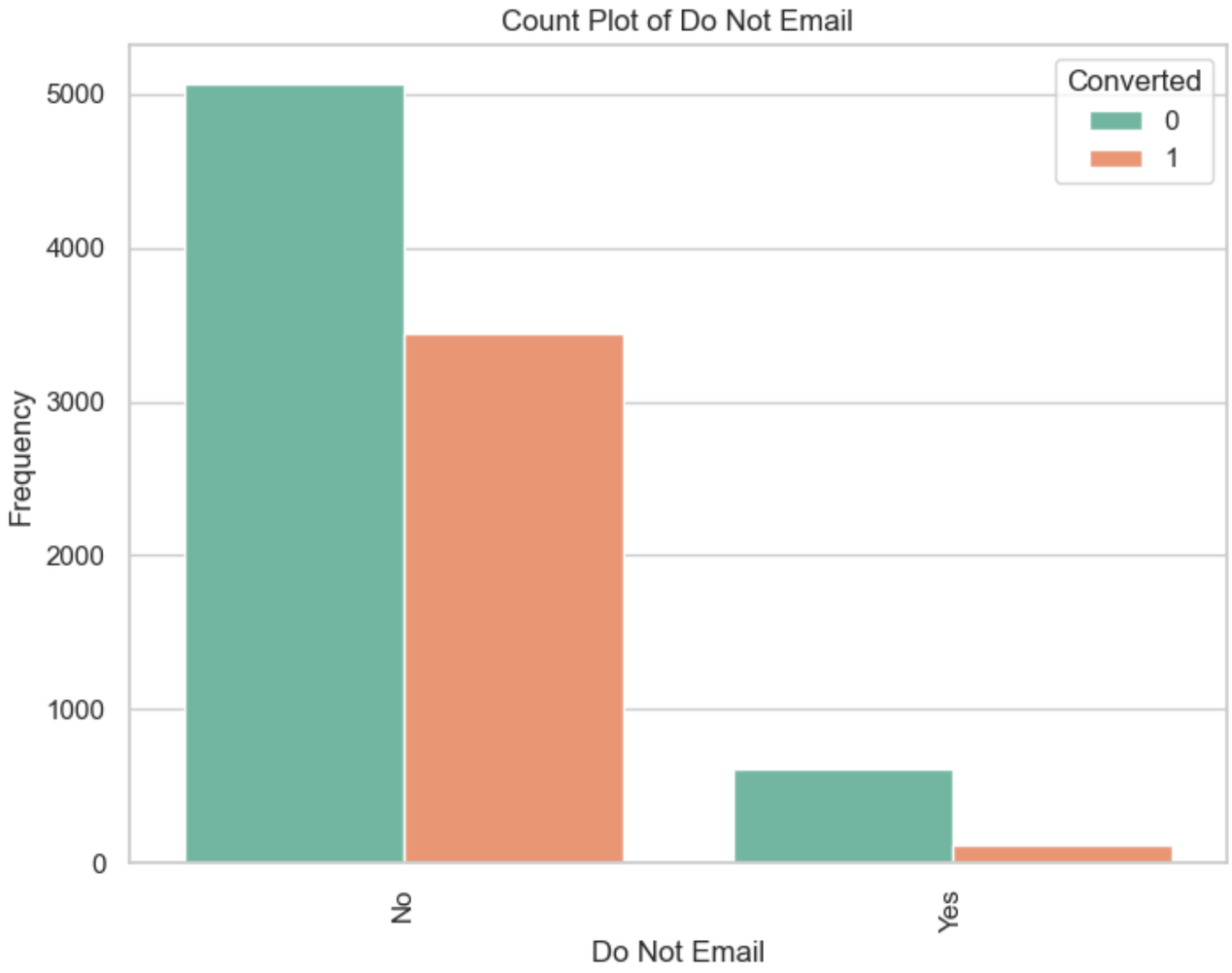
2. Google-sourced leads exhibit the highest probability of conversion.

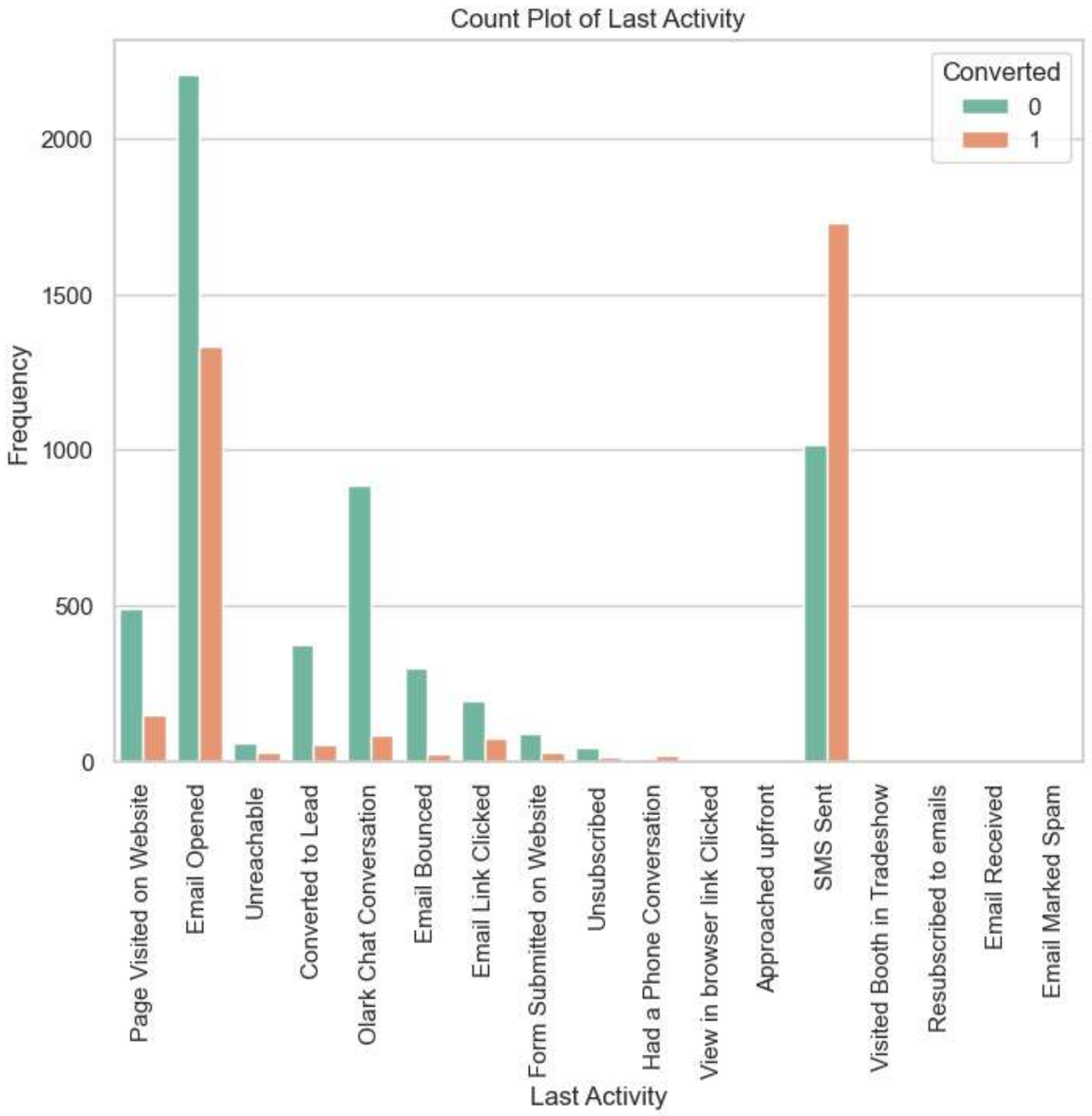
3. Leads from the Reference and Welingak Website source have a notable probability of conversion.

Do Not Email

1. Customers opting for Do Not Mail show a lower conversion rate.

2. Majority of leads are from customers who do not opt for Do Not Mail, with a higher conversion rate.





Last Activity

1. Customers whose last activity was SMS Sent show a higher conversion rate .

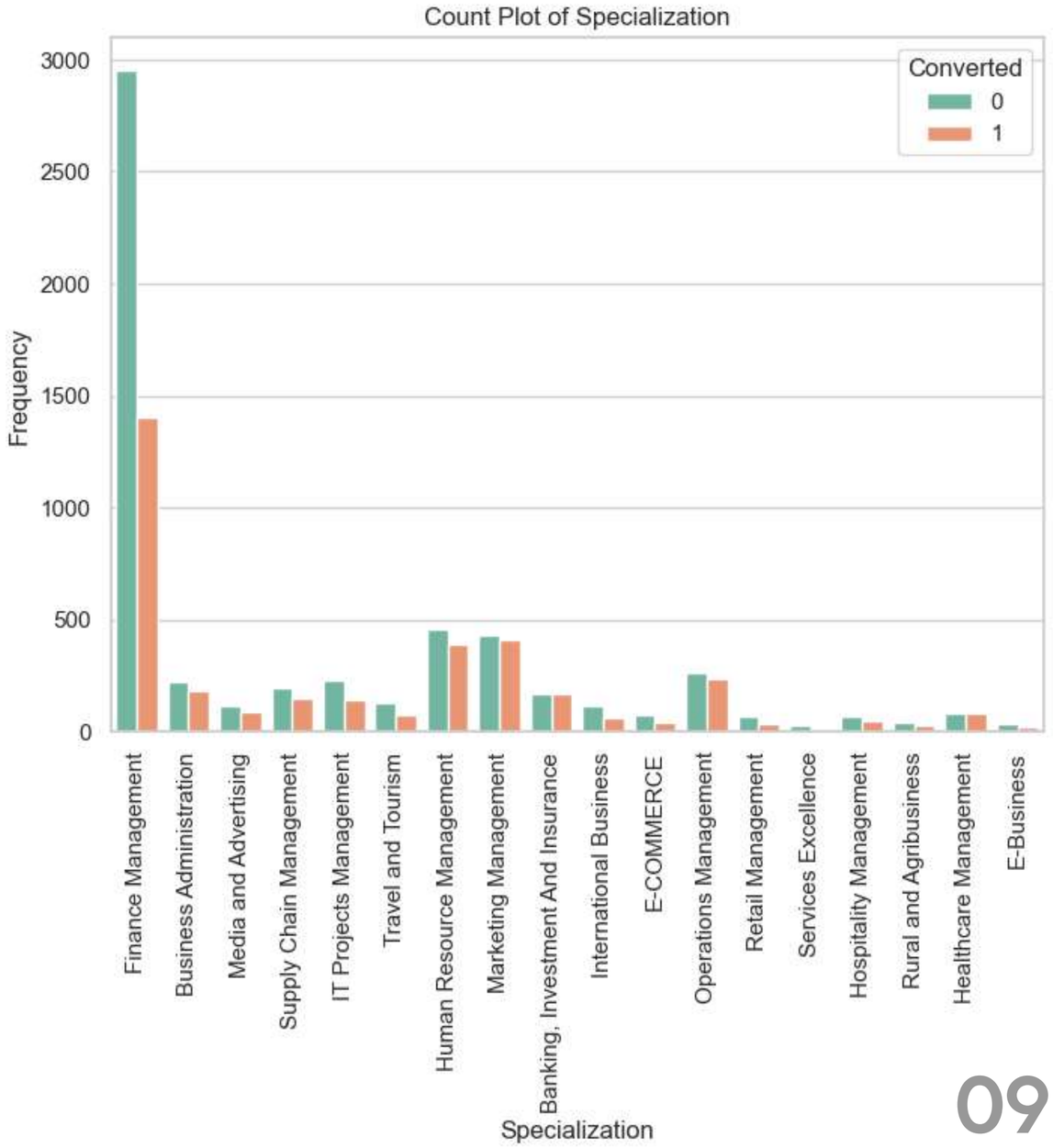
2. Customers with the last activity as Email Opened constitute the majority, with a conversion rate .

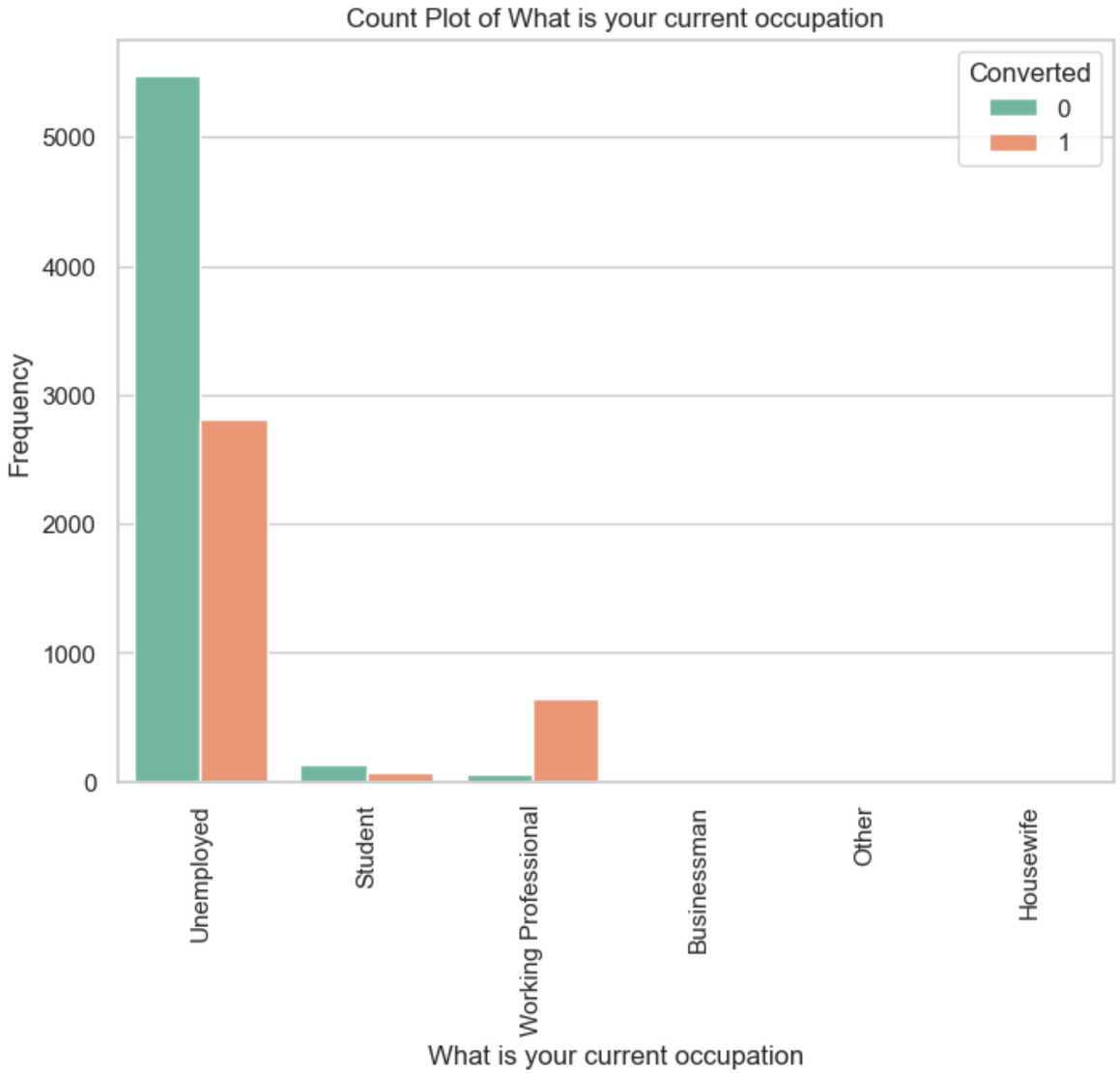
To improve the overall conversion rate, focus on improving conversions from customers with the last activity as Email Opened and generate more leads from those with the last activity as SMS Sent.

Specialization

1. The majority of leads have specializations in Management & Others.

2. Leads specialized in Rural & Agribusiness show the least probability of conversion.

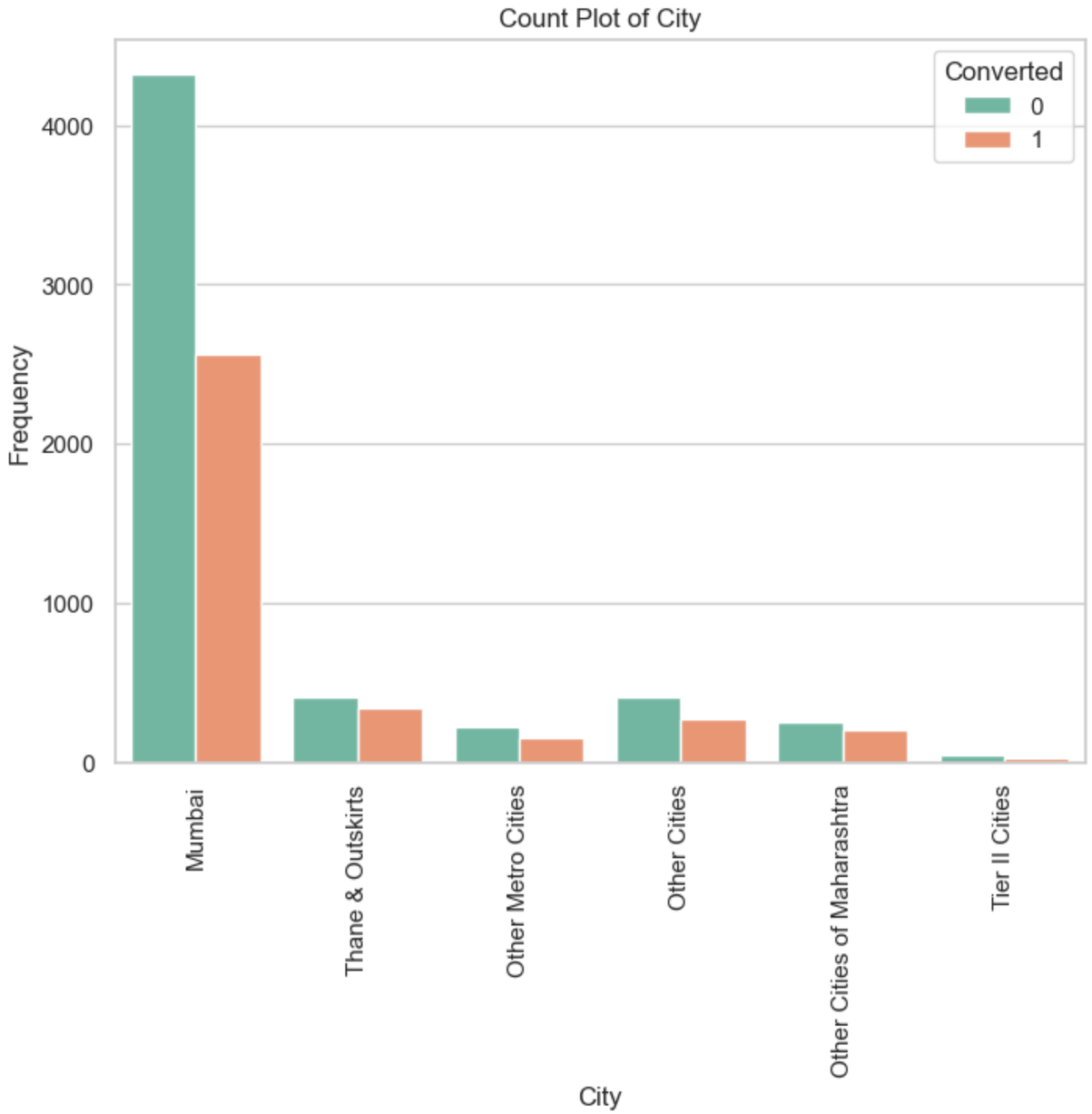




Current Occupation

1. Maximum leads are Unemployed.

2. Working Professional are the one who have more then 90% of conversion



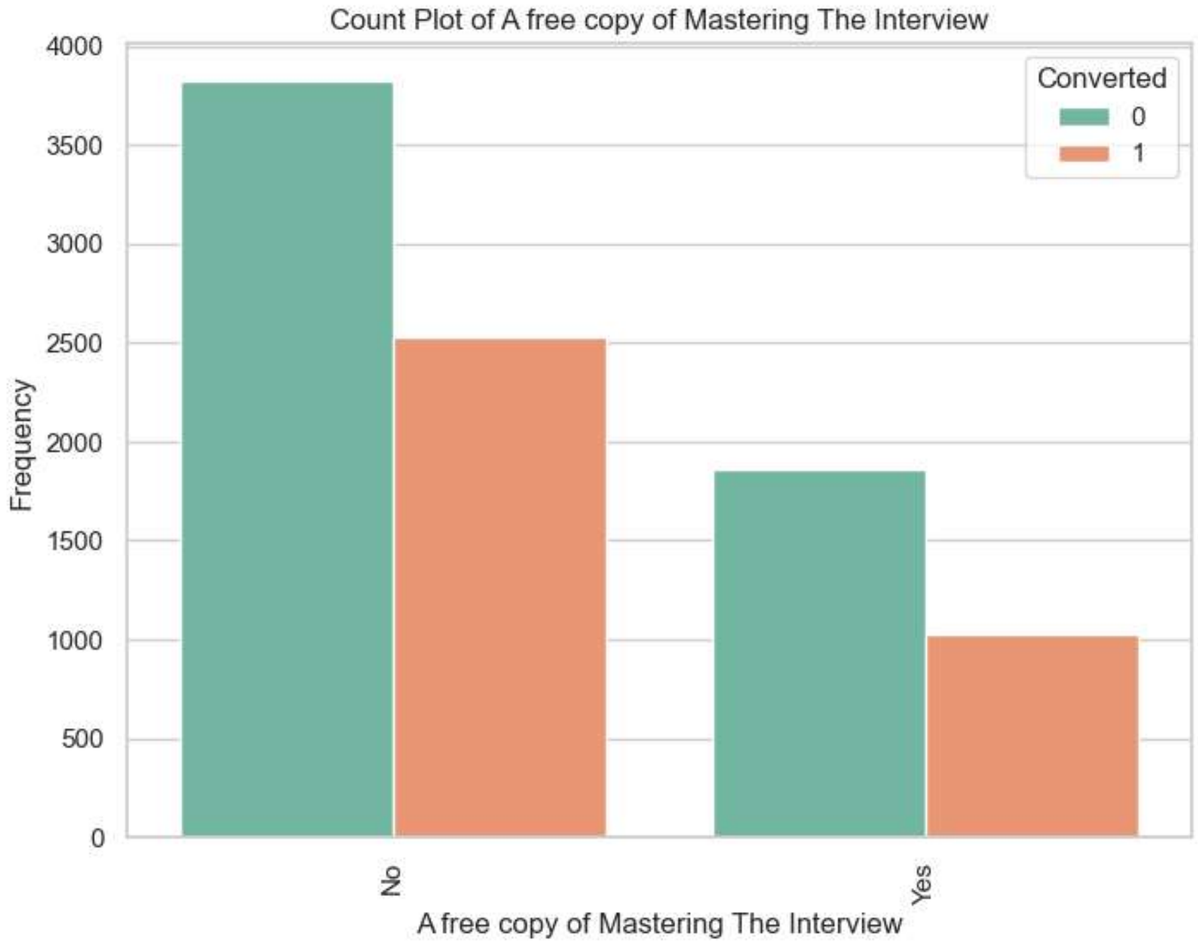
City

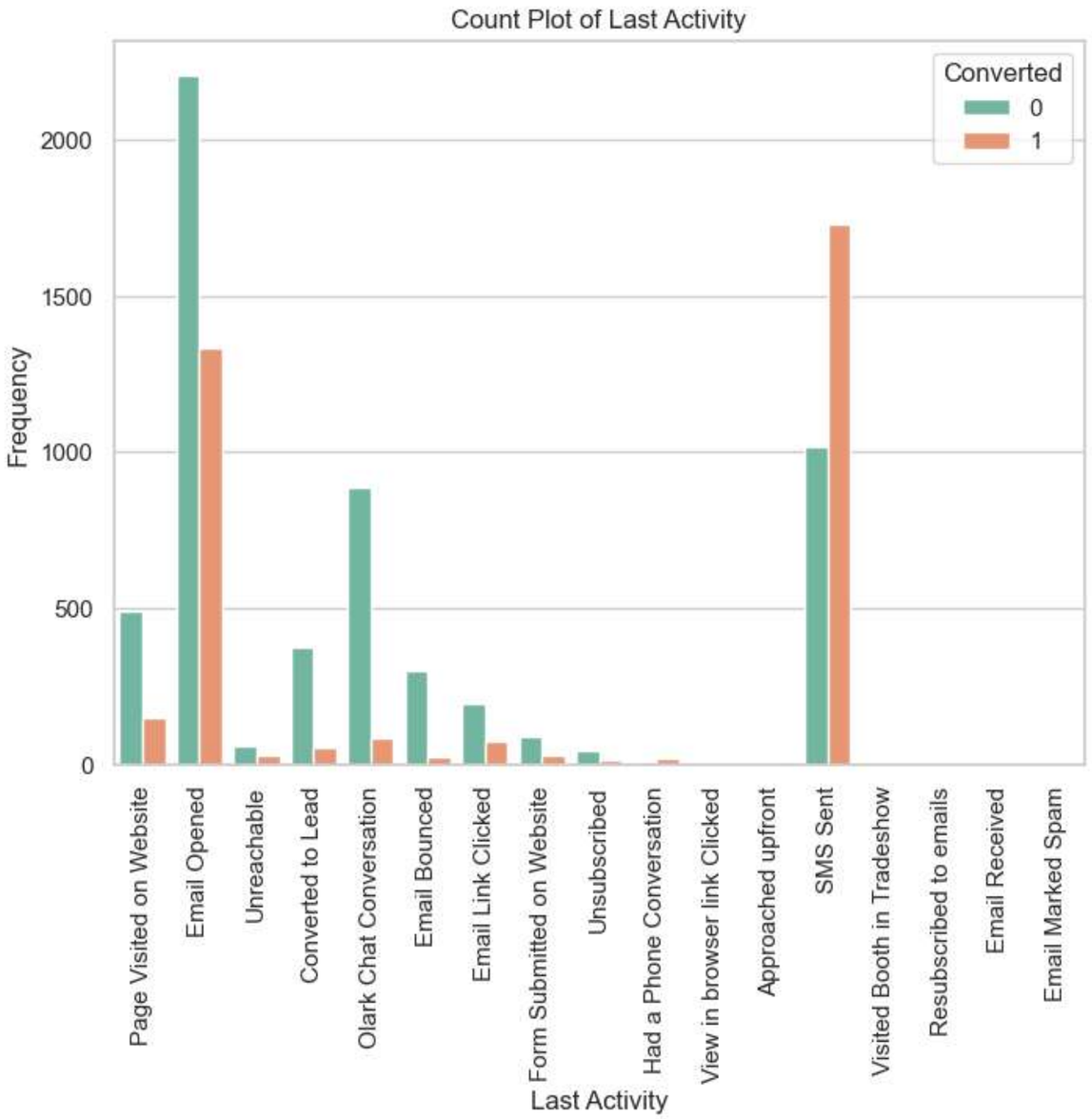
- The most leads are from Mumbai, but the highest conversion rate is in Thane & Outskirts.
- The conversion rate for Other Metro Cities is higher than the overall conversion rate, suggesting that these cities may be a good place to focus your marketing efforts.
- The conversion rate for Other Cities and Other Cities of Maharashtra is lower than the overall conversion rate.
- The conversion rate for Tier II Cities is the lowest, suggesting that this may be a more challenging market to convert leads.

A free copy of Mastering The Interview

- Leads who were offered a free copy of Mastering the Interview were more likely to convert than those who were not. The conversion rate for leads who received a free copy of the book was 35%, while the conversion rate for leads who did not was only 20%.

- This suggests that offering leads a free copy of Mastering the Interview can be a very effective way to increase conversion rates.





Last Notable Activity

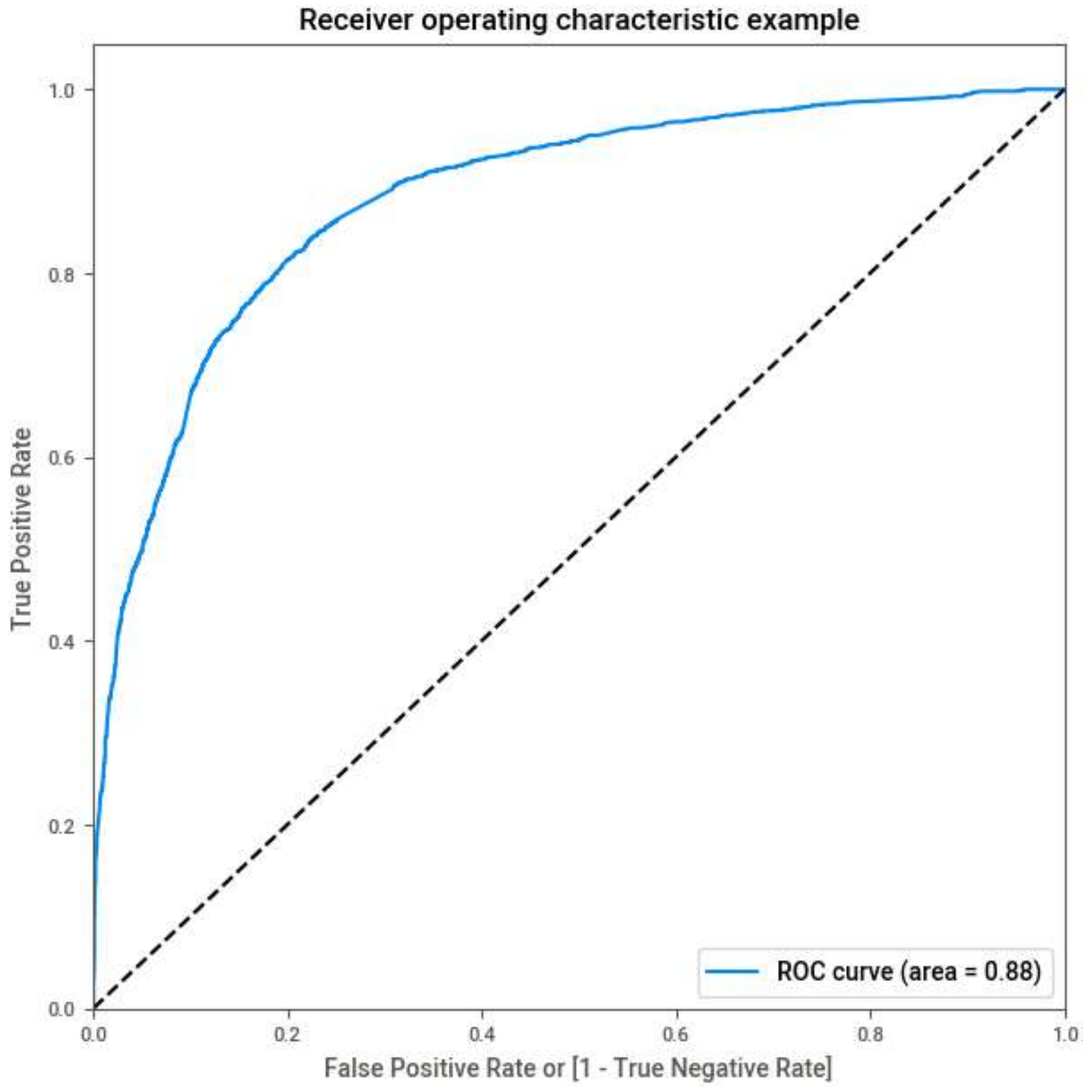
- A significant number of converted leads have "Modified" their contact information. This suggests that it is important to keep your lead database up to date, as converted leads are more likely to change their contact information than leads who have not yet converted.

- A relatively large number of converted leads have "Unreachable" or "Unsubscribed" email addresses. This suggests that it is important to clean your email list on a regular basis to remove invalid email addresses and unsubscribes.

- A relatively small number of converted leads have "Had a Phone Conversation" or "Olark Chat Conversation". This suggests that these channels may not be as effective at converting leads as other channels, such as website visits or email.

ROC Curve

- In this case, the ROC curve exhibits a high degree of accuracy, with an area under the curve (AUC) value of 0.88. This indicates that the classifier is capable of correctly identifying a majority of positive cases while minimizing the number of false positives.



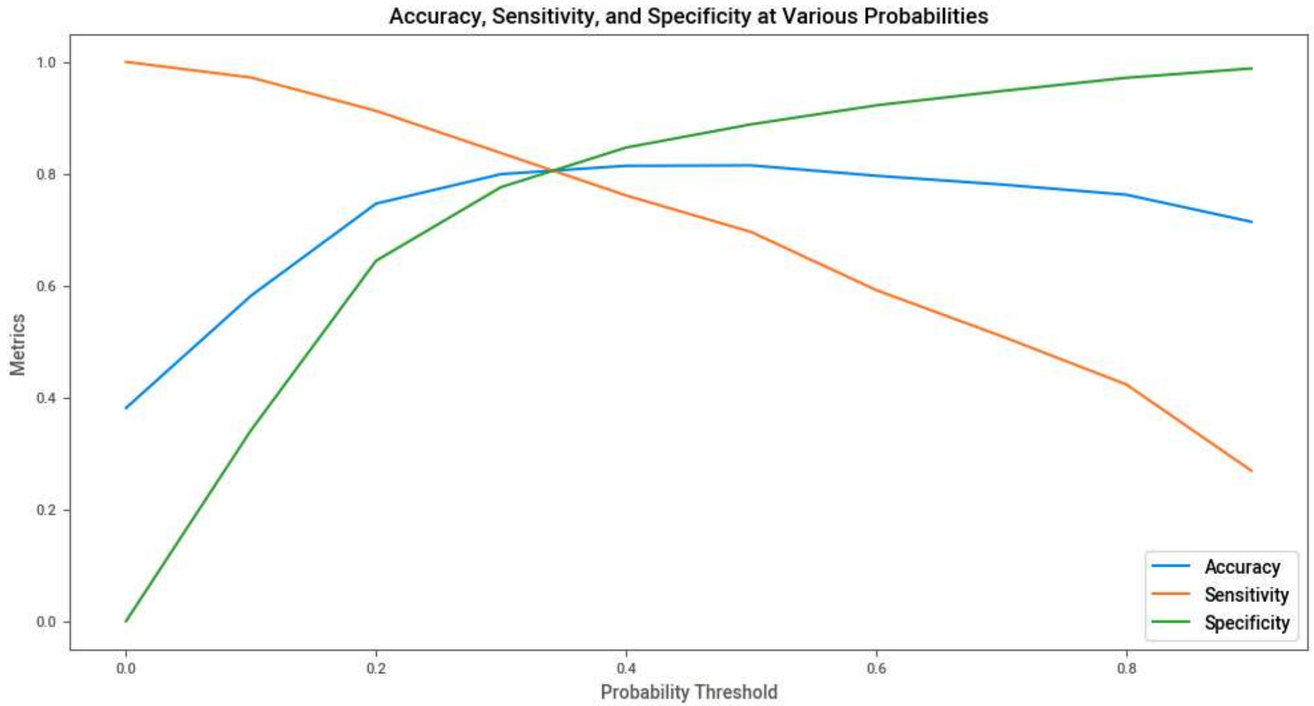
Accuracy, Sensitivity, and Specificity at Various Probabilities

•The graph shows the relationship between accuracy, sensitivity, and specificity at various probabilities.

•Accuracy is the proportion of all predictions that are correct, regardless of whether they are positive or negative.

•Sensitivity is the proportion of positive cases that are correctly identified.

•Specificity is the proportion of negative cases that are correctly identified.



Model Evaluation

•The model performs slightly better on the test data than on the training data, which suggests that it is generalizing well to new data.

•The model has a precision of 81%, which means that 81% of the time it predicts a positive case, it is correct.

•The model has a recall of 81%, which means that it identifies 81% of all true positive cases.

•The model has an F1-score of 81%, which is a harmonic mean of precision and recall. This suggests that the model has a good balance between precision and recall.

Overall, the model performs well on both the training and test data. It is able to accurately identify both positive and negative cases.

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

Insights from the Lead Conversion Data
In conclusion, the data presented in this presentation provides valuable insights into the behavior of leads and the factors that influence conversion rates, such as city, tag, and notable activity. The insights gained from this data can be used to develop targeted interventions to improve conversion rates.

THANK YOU