

LEAD CASE STUDY

Debdatta Sarkar
26th May 2023



OUTLINE

01 EXECUTIVE SUMMARY

Summarizes the process and report

02 INTRODUCTION

03 METHODOLOGY

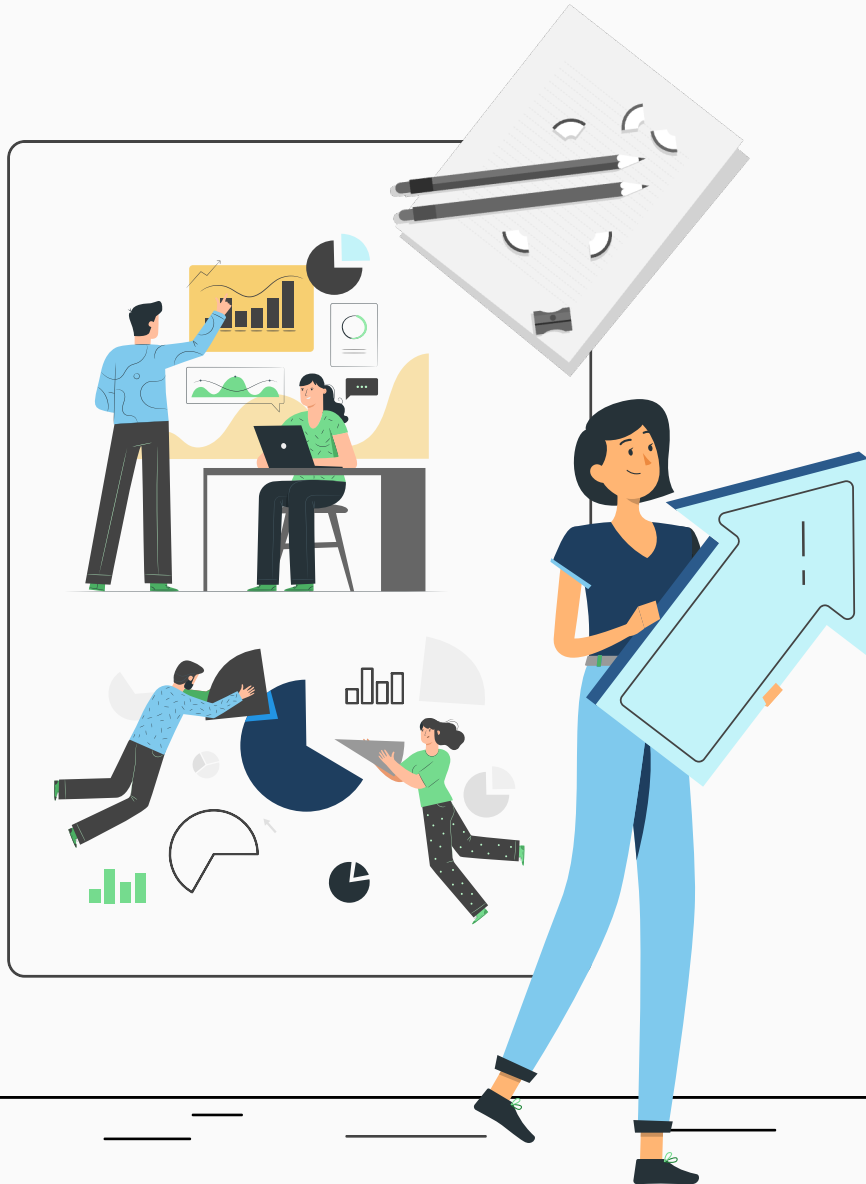
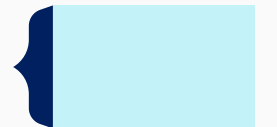
Overview of goals and objectives and the purpose of the report
Process followed for the goals to be reached

04 RESULTS

Consequence of EDA and visualizations

05 CONCLUSION

Condensed summary of ideas, viewpoints and persuasive arguments



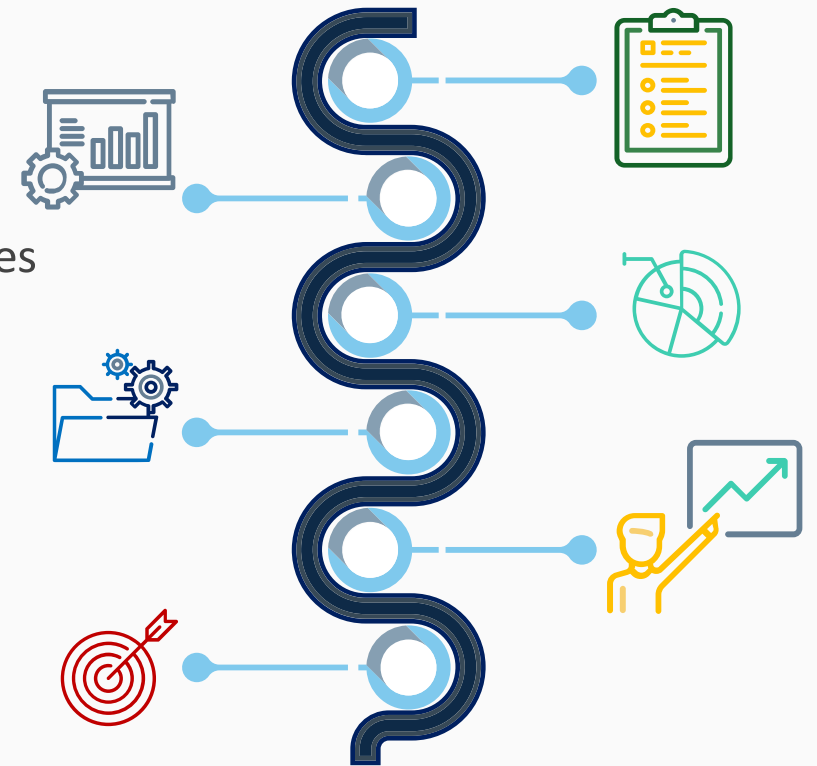
OI EXECUTIVE SUMMARY

■ SUMMARY OF METHODOLOGIES -

- Data Collection
- Quality Checks
- Exploratory Data Analysis
- Identifying Categorical Variables and Creating Dummy Variables
- Model Building Using Logistic Regression
- Predictions On Test Dataset
- Conclusion

■ SUMMARY OF ALL RESULTS -

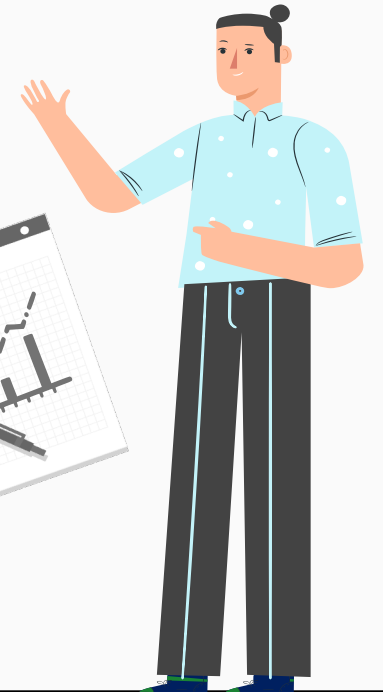
- Data Analysis along with Interactive Visualizations
- Conclusion and recommendations



02 INTRODUCTION

PROJECT BACKGROUND AND CONTEXT

X Education is an online education company that specializes in offering courses to professionals in various industries. The company acquires leads through multiple channels, including website visits, form submissions, and referrals. However, their current lead conversion rate stands at a relatively low 30%. In order to enhance their operational efficiency, the company aims to identify the most promising leads, referred to as "Hot Leads," and concentrate their sales efforts on them. Their objective is to develop a model that assigns a lead score to each potential customer, indicating the likelihood of their conversion. To meet this objective, the CEO has set a target lead conversion rate of 80%. The company has provided a dataset containing historical lead information, encompassing attributes such as Lead Source, Total Time Spent on Website, and Last Activity. The primary objective of this case study is to construct a logistic regression model that assigns lead scores and addresses any potential future issues outlined in a separate document.



PROBLEMS WE WANT TO FIND ANSWERS & PROCESSES

- Exploring prospective leads
- Developing a model for lead scoring
- Enhancing the rate of lead conversion
- Managing categorical variables effectively
- Ensuring future adaptability
- Providing recommendations and delivering a presentation

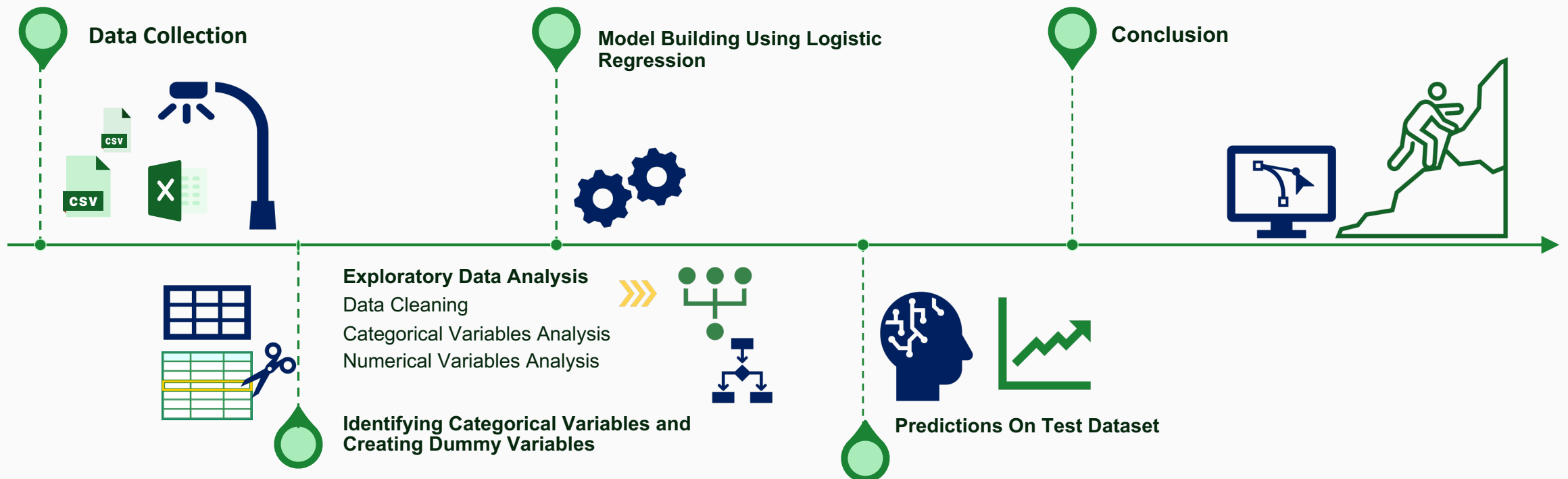
03. METHODOLOGY

Process followed for the goals to
be reached



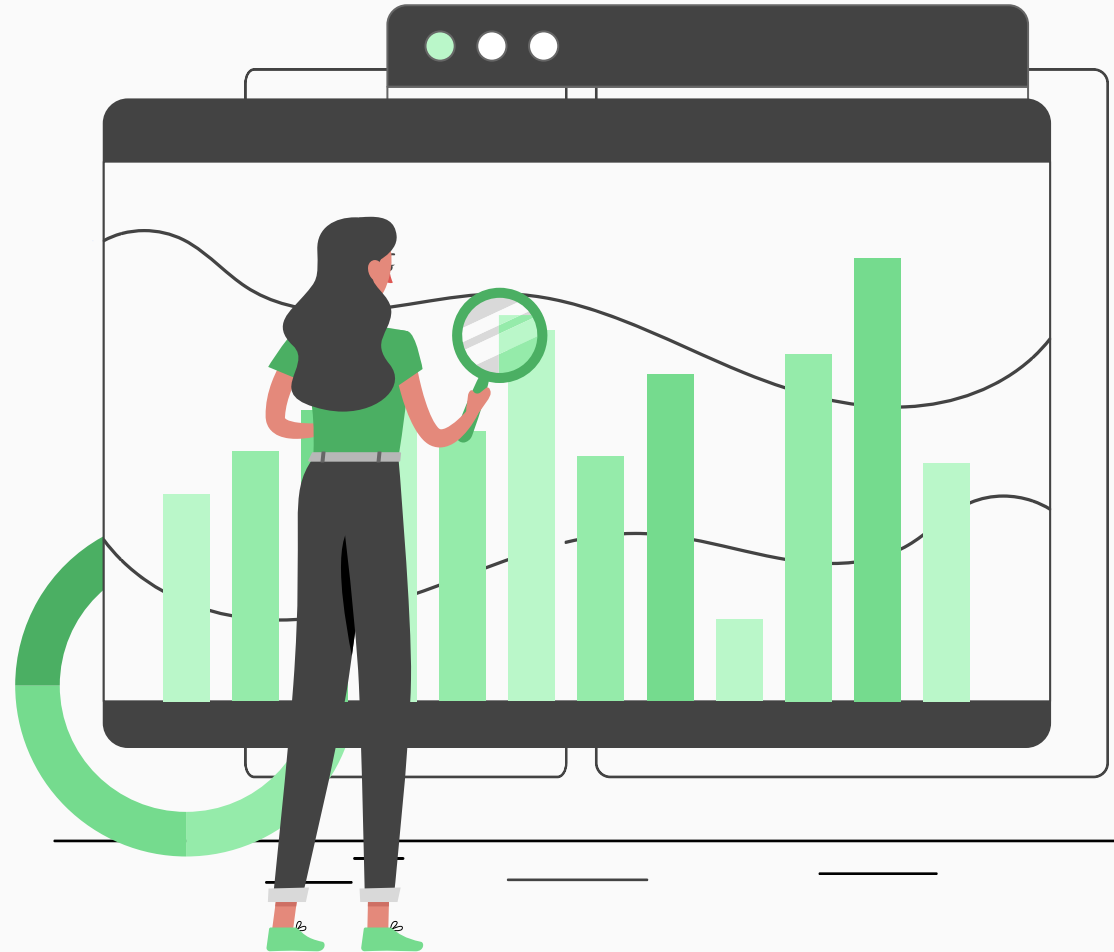
SUMMARY OF STEPS

The complete process involves gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate conclusion and outcomes.



04. RESULTS

Consequence of EDA and
visualizations



EDA WITH VISUALIZATION

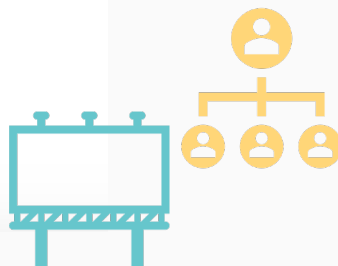
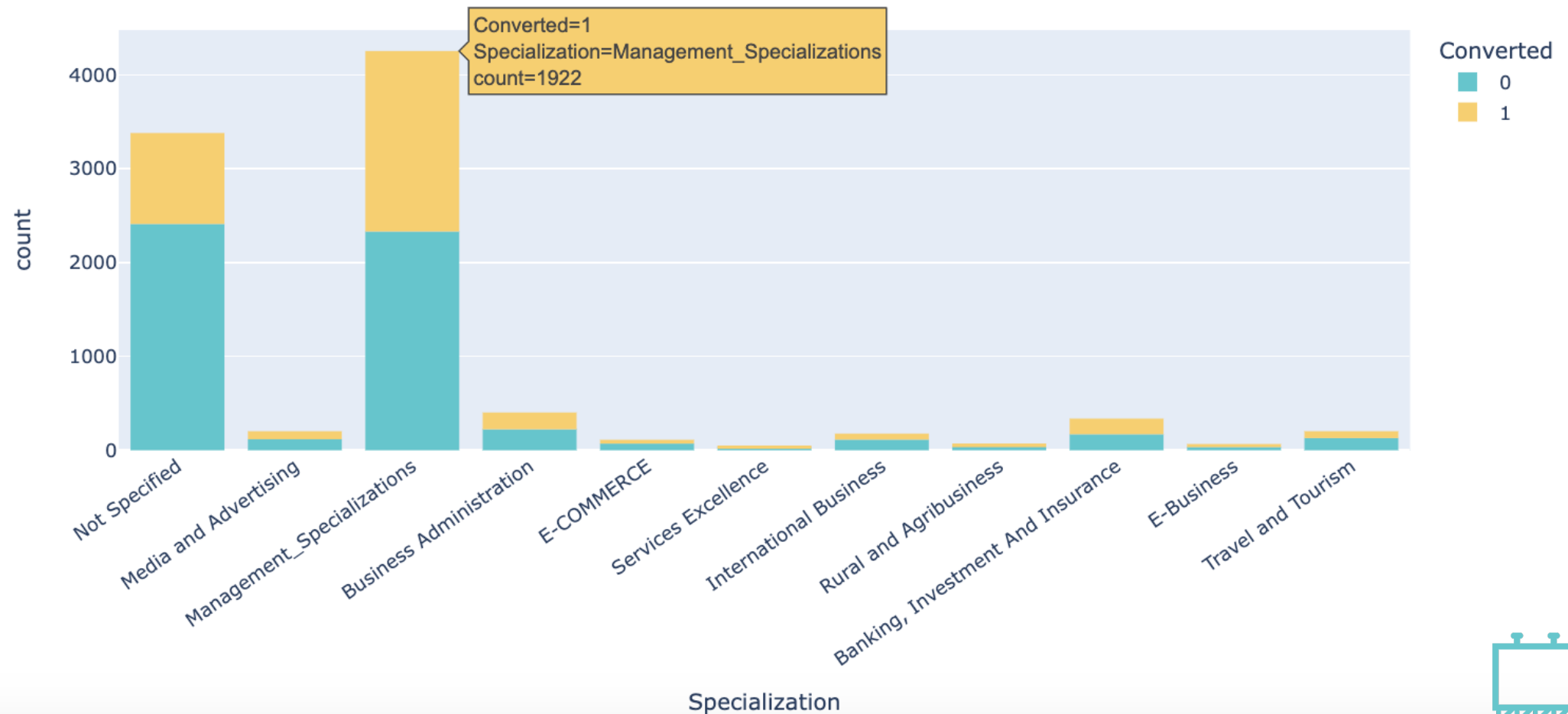
Analysis via visual graphs



SPECIALIZATION ANALYSIS



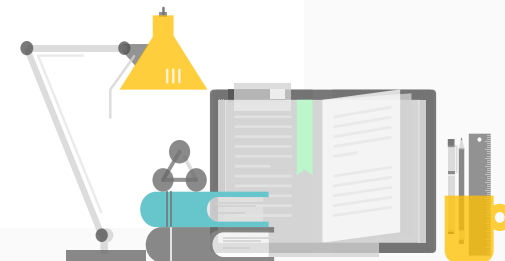
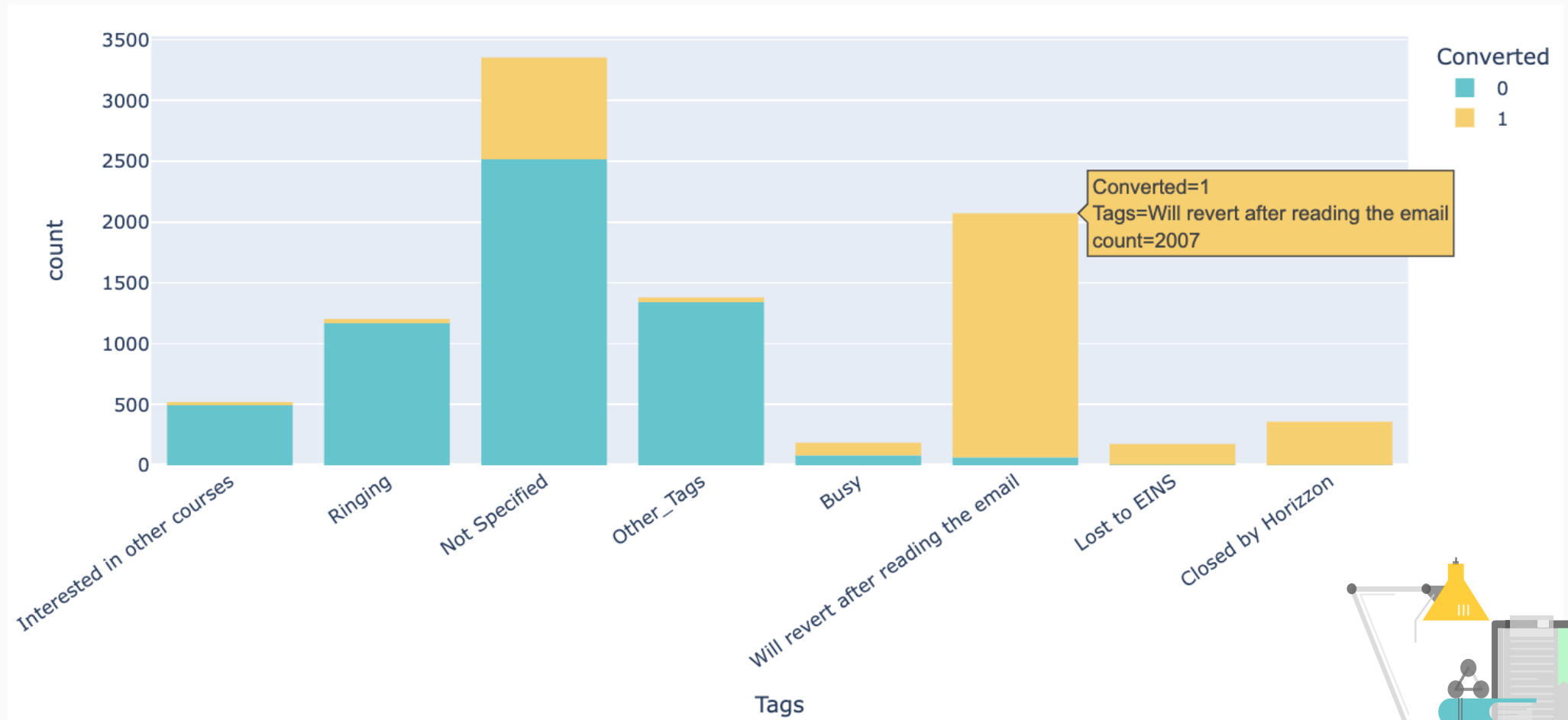
Management Specialization has maximum conversation count 1922.



COMMUNICATIONS & MEDIA ANALYSIS



Email communication seems to have large conversation rate with count if 2007.



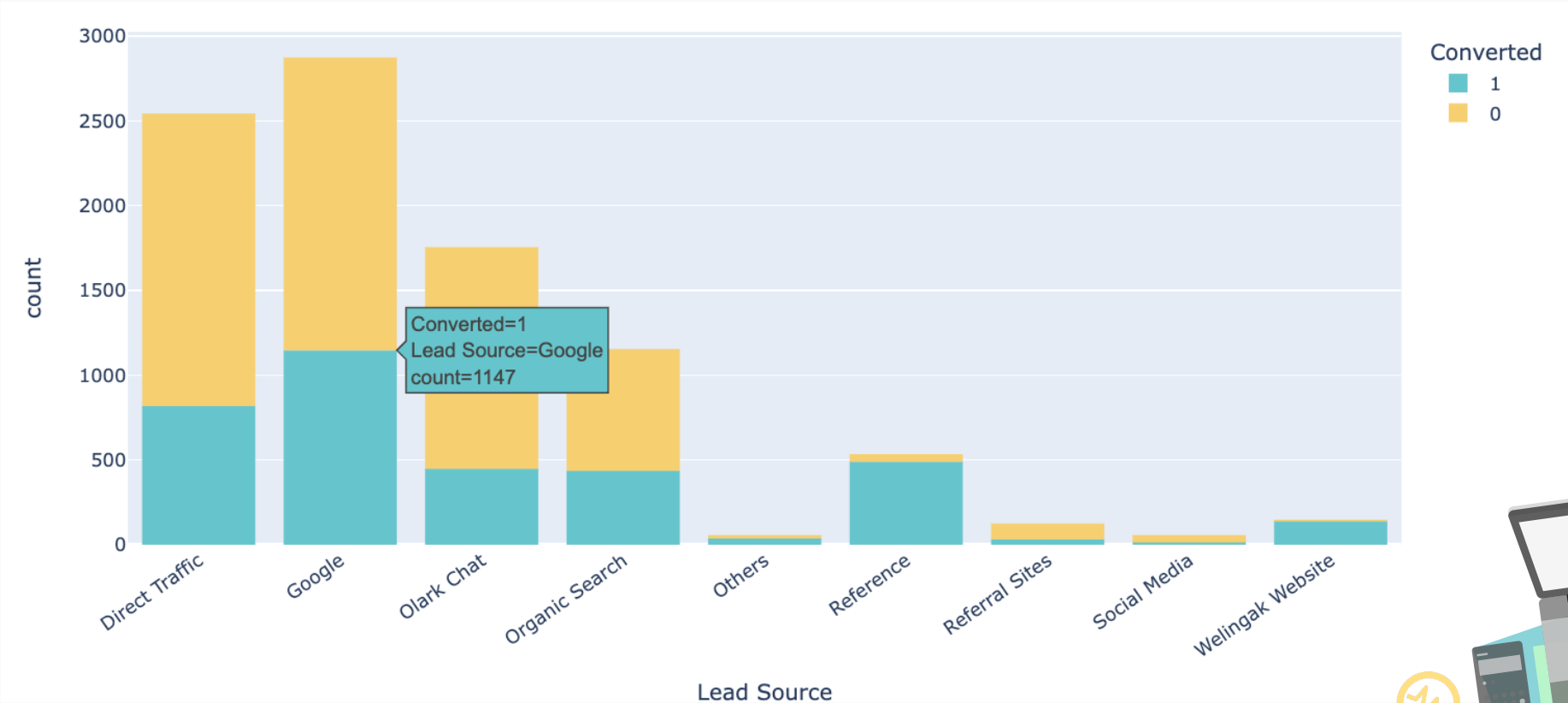
LEAD SOURCE ANALYSIS



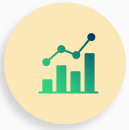
The column containing data for 'Google' and 'Direct Traffic' shows the highest number of leads and lead conversions, while 'Welingak Website', 'Reference', and 'Others' exhibit the highest conversion rates among leads. Given its importance, it is crucial to retain this column.



To enhance the overall lead conversion rate, it is recommended to concentrate efforts on improving the conversion of leads from 'direct traffic' and 'Google'. Additionally, it is crucial to generate more leads from 'reference' and 'Welingak website' as they demonstrate a strong conversion rate. By prioritizing these areas, the overall lead conversion rate can be significantly improved.



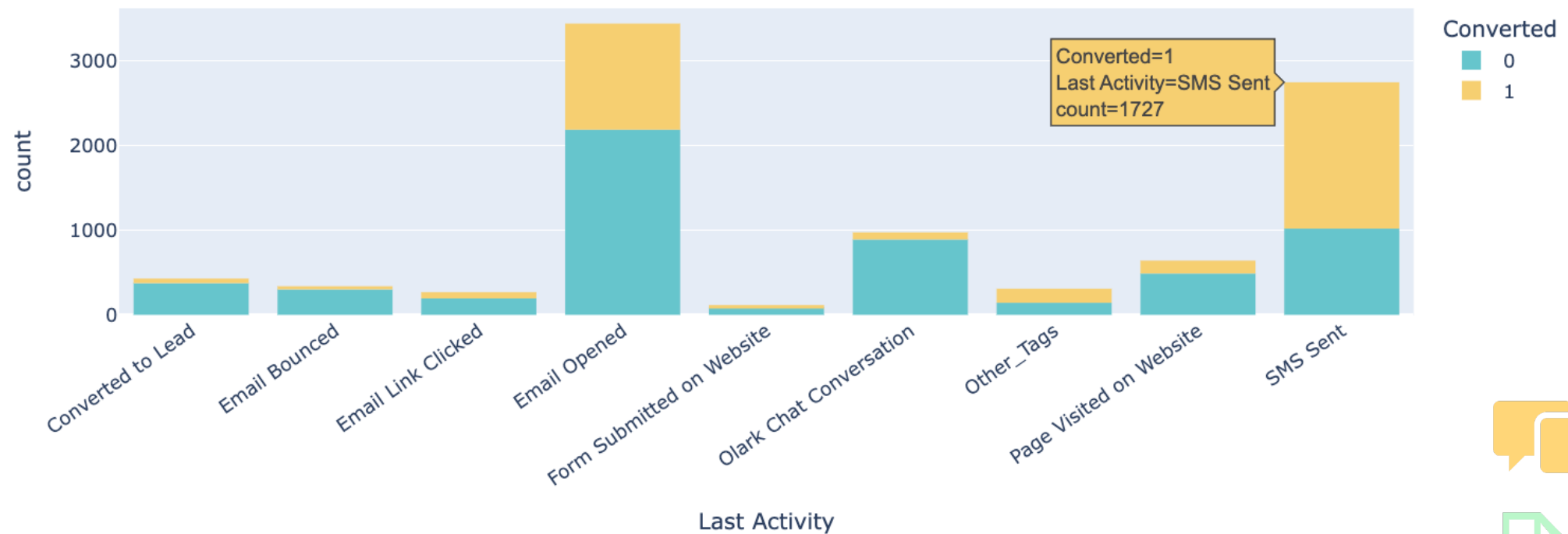
LAST ACTIVITY ANALYSIS



Leads with the last activity recorded as 'SMS Sent' display the highest conversion rate. 'Email Opened' generates the largest quantity of leads and ranks second in terms of conversion rate.



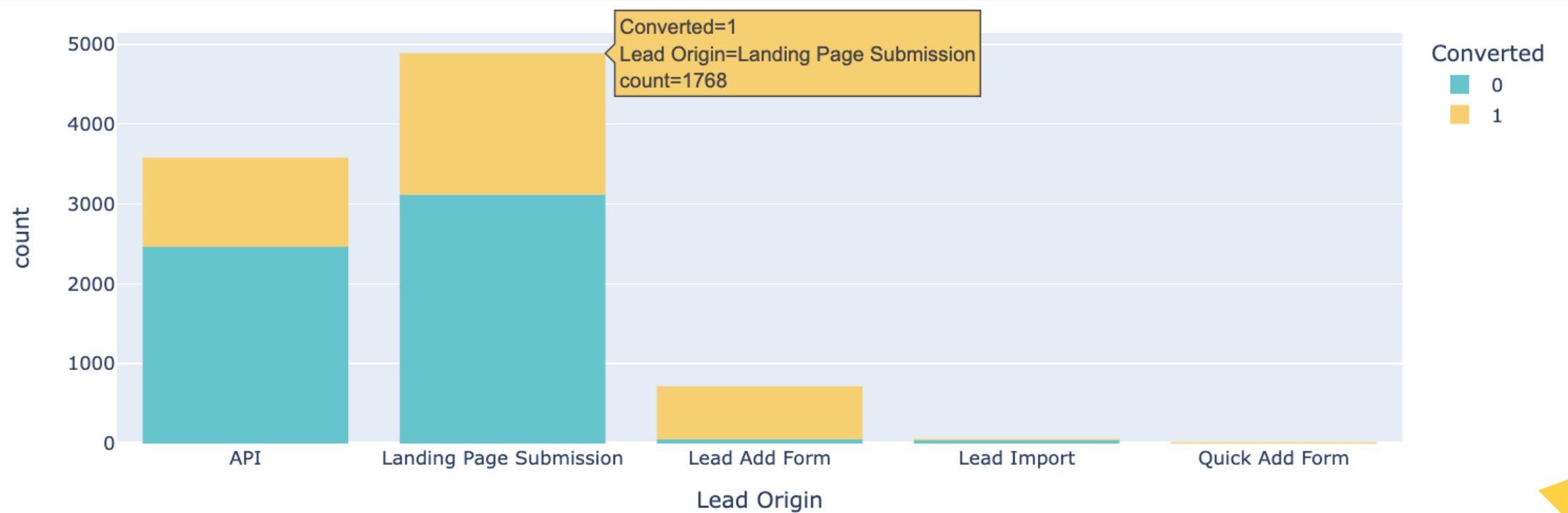
To optimize results, it is advisable to prioritize leads with a last activity of 'SMS Sent' due to their exceptional conversion ratios.



LEAD ORIGIN ANALYSIS



'Landing Page Submissions' have a high lead conversion ratio with a sufficient volume to increase the number of converted leads. To improve the overall lead conversion rate, it is crucial to enhance the lead conversion from the API and Landing Page Submission sources. Additionally, efforts should be focused on generating more leads from the Lead Add Form.



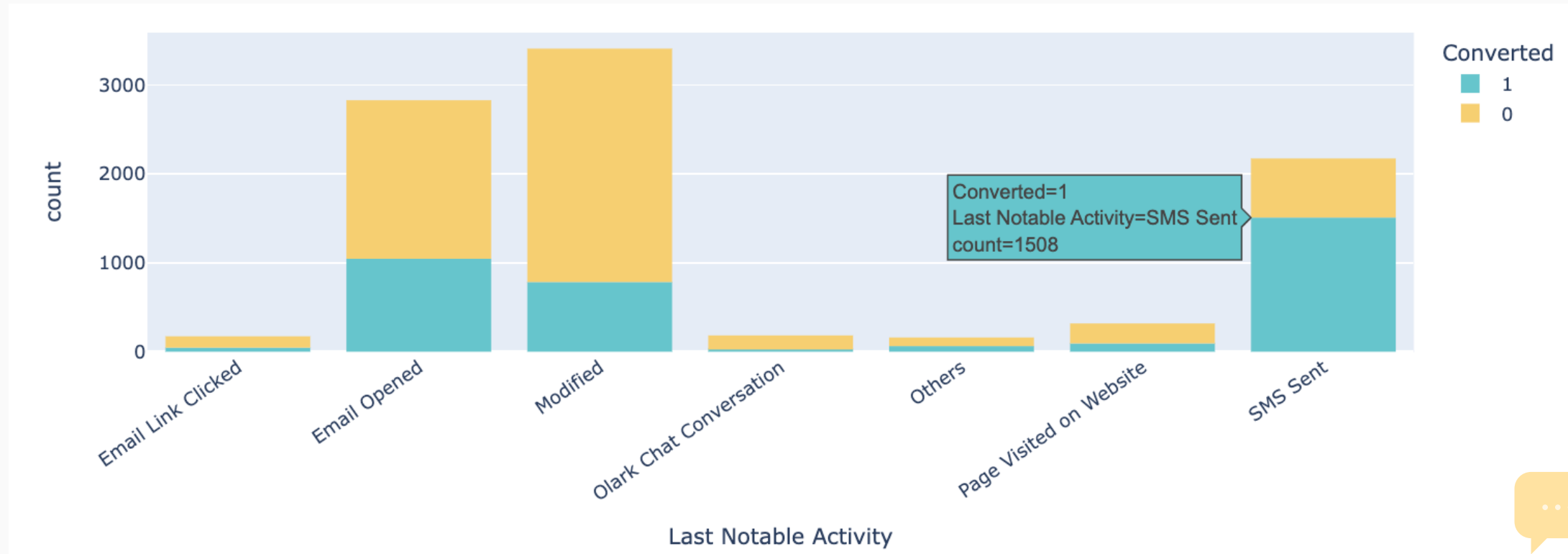
LAST NOTABLE ACTIVITY ANALYSIS



The leads with the last activity recorded as 'SMS Sent' show the highest conversion rate. 'Email Opened' generates the maximum number of leads and also ranks second in terms of conversion rate.



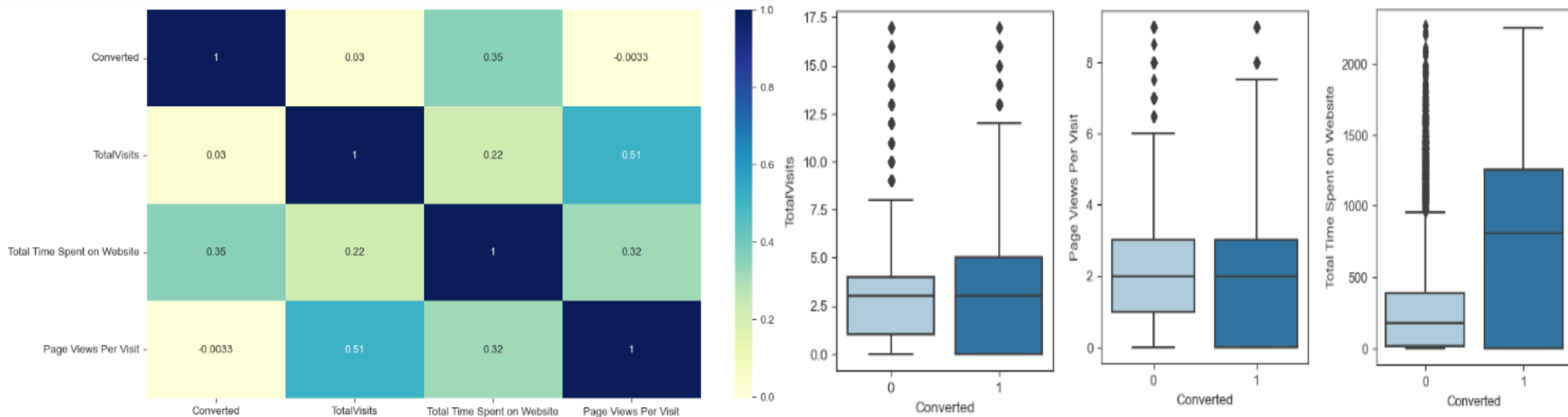
To enhance overall conversion rates, it is advisable to prioritize leads where the last activity is 'SMS Sent' as they consistently demonstrate the highest conversion ratios.



BIVARIATE ANALYSIS – HEATMAP & BOXPLOTS

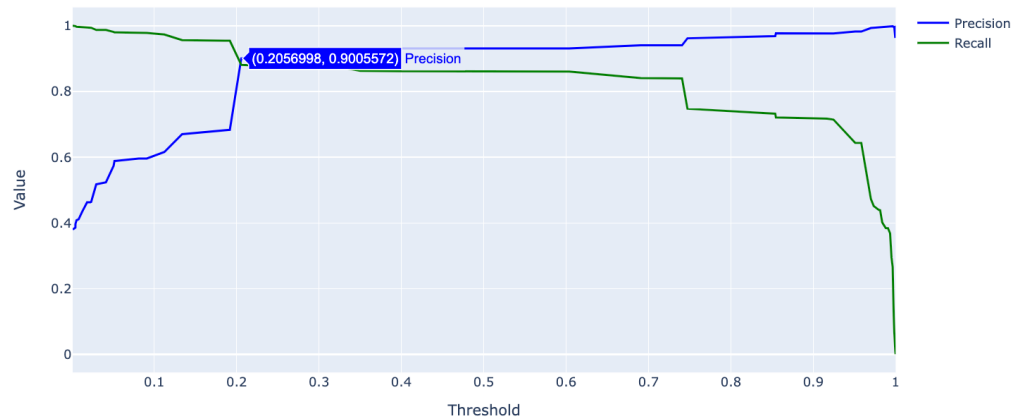


- Customers who spends more time on the Website have a higher chance of getting successfully converted.
- Same appears on heatmap the correlation between total time spent w.r.t. to conversation is high i.e., 0.36.

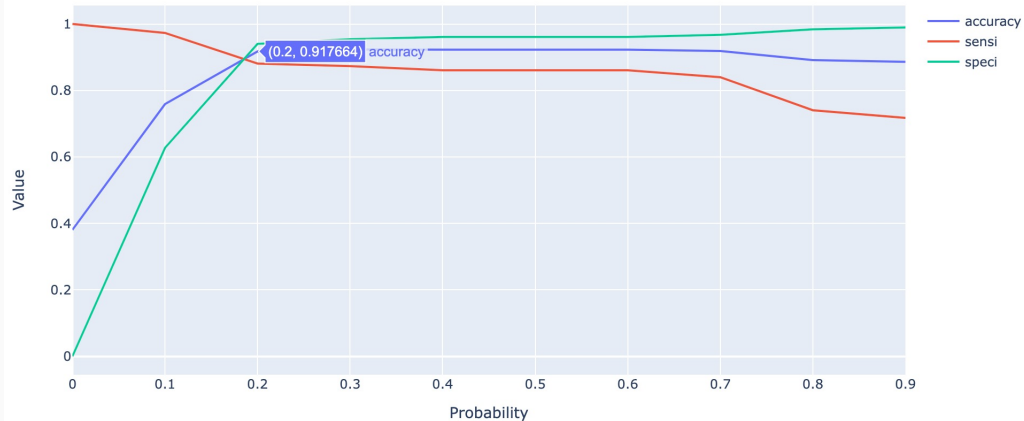


ROC CURVE

Precision and Recall vs. Thresholds



Accuracy, Sensitivity, and Specificity



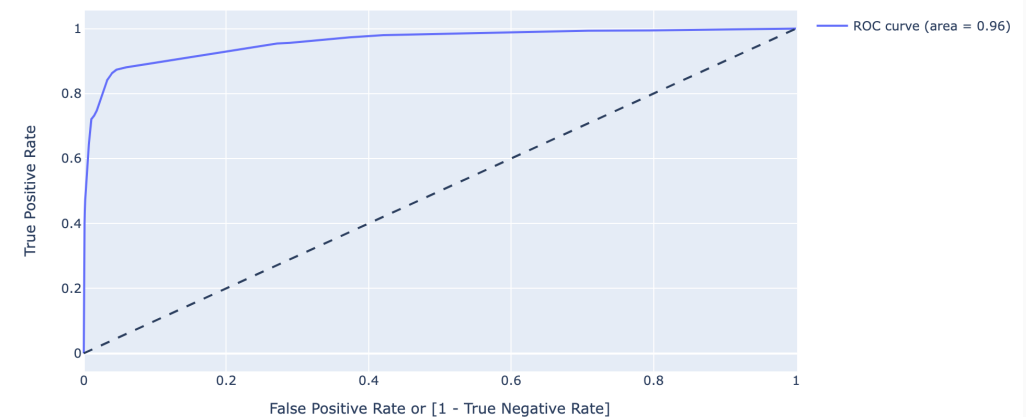
Another tool to measure the classification model performance visually is ROC Curve. **ROC Curve** stands for **Receiver Operating Characteristic** Curve. An ROC Curve is a plot which shows the performance of a classification model at various classification threshold levels.

The ROC Curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels.

True Positive Rate (TPR) is also called Recall. It is defined as the ratio of TP to (TP + FN).

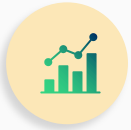
False Positive Rate (FPR) is defined as the ratio of FP to (FP + TN).

Receiver Operating Characteristic (ROC) Curve



ROC curve should be a value closer to 1 for a good model. We have got a value of 0.96 which is extremely good.

MODEL EVALUATION



The cut-off value is near near to 0.2. Sensitivity is 88.09% of train set and 88.71% of test dataset. CEO of X-Education has target of 70% and the model has achieved accuracy of ~91%.

Training Data Confusion Matrix			
	Predicted		
Actual		Negative	Positive
	Negative	TN = 3650	FP = 232
	Positive	FN = 284	TP = 2101

Test Data Confusion Matrix			
	Predicted		
Actual		Negative	Positive
	Negative	TN = 1582	FP = 94
	Positive	FN = 114	TP = 896

	Train Dataset	Test Dataset		Train Dataset	Test Dataset
Accuracy	91.77 %	92.26 %	Precision	90.06 %	90.51 %
Sensitivity	88.09 %	88.71 %	Recall	88.09 %	88.71 %
Specificity	94.02 %	94.39 %			

RECOMMENDATION BASED ON FINAL MODEL

- Media communications like Email, SMS, forms and Chat conversations should be more encouraged. (Lead Origin_Lead Add Form, Lead Source_Direct Traffic)
- Allocating a larger budget or increasing spending on advertising and other initiatives for the Welingak Website would be beneficial. (Lead Source_Welingak Website)
- Proper status needs to be maintained for tags and customers need to be followed up and the whole thing can be automated with modern tools. (Tags_Closed by Horizon, Tags_Busy, Tags_Lost to EINS, Tags_Will revert after reading the email, Tags_Ringing, Tags_Not Specified)
- Customers last activity needs to be tracked as well and accordingly can be pursued. (Last Activity_Email Bounced, Last Activity_Olark Chat Conversation, Last Notable Activity_SMS Sent)

	Features	VIF
2	Lead Source_Welingak Website	1.35
6	Tags_Closed by Horizon	1.15
3	Last Activity_Email Bounced	1.05
5	Tags_Busy	1.04
7	Tags_Lost to EINS	1.0
0	Lead Origin_Lead Add Form	0
4	Last Activity_Olark Chat Conversation	0
11	Last Notable Activity_SMS Sent	0
1	Lead Source_Direct Traffic	0
10	Tags_Will revert after reading the email	0
9	Tags_Ringing	0
8	Tags_Not Specified	0





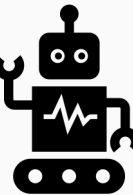
05 CONCLUSION

1.Segment leads based on responsiveness: Working professionals who are categorized as management professionals brought in more conversion. Categorize the potential leads based on their responsiveness to previous communication attempts. Prioritize those who have shown interest or engagement by responding to emails, texts, or any other communication channels. Lead which came through sources such as welingak website and direct traffic. Lead who are spending lot of time on website and are coming back frequently.

2.Prioritize potential leads: Identify and prioritize the potential leads who have been predicted as "1" by the model. These individuals have a higher likelihood of converting, so they should be the primary focus during this period.

3.Utilize automation tools: Utilize automation tools, such as CRM (Customer Relationship Management) software, to streamline and track the outreach process. These tools can assist in managing lead data, scheduling calls, and monitoring follow-ups, ensuring efficient and consistent communication with potential leads.

4.Implement personalized outreach: Develop personalized scripts or templates for phone calls, tailored to address the specific needs and pain points of each potential lead.



THANKS

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik** and illustrations by **Storyset**

Please keep this slide for attribution.

