LEAD SCORE CASE STUDY

BY:

SONALI GHEDIYA

SONIA VIJAYKUMAR

MUKUNTH S

GOAL

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

BUSINESS OBJECTIVE

- X Education seeks to identify the most promising leads to optimize their sales efforts.
- They aim to develop a predictive model capable of identifying hot leads effectively.
- The deployment of this model is essential for future use and maximizing lead conversion.

PROCESS FOLLOWED

Loading & Observing the past data provided by the Company

Univariate, Bivariate, and Heatmap for numerical and categorical columns

Performing pre-requisites for RFE and Logistic Regression

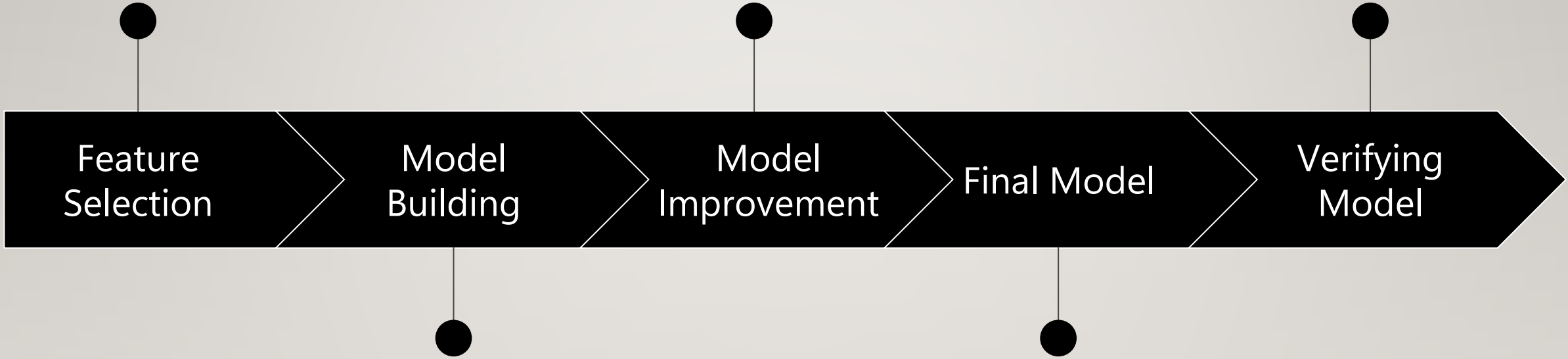
Duplicate removal, null value treatment, unnecessary column elimination, etc.

Outlier Treatment, Feature-Standardization

Selection of top 20 features using RFE

Reduction of columns and Model re-building

Verifying our Final Model Accuracy etc.

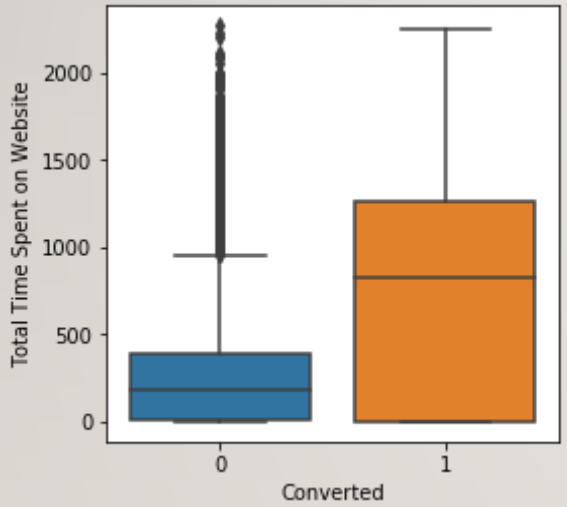


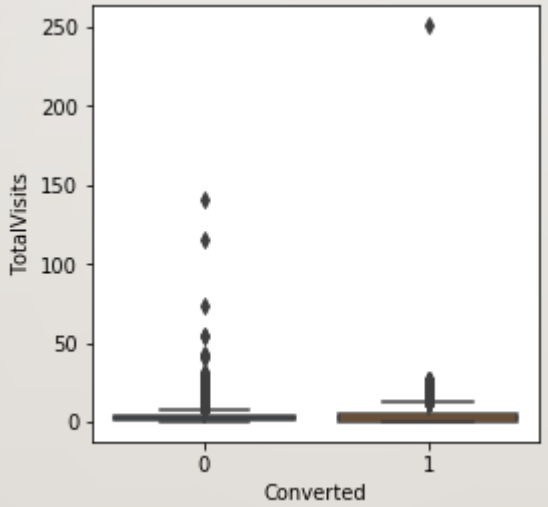
Model building using RFE for selected columns

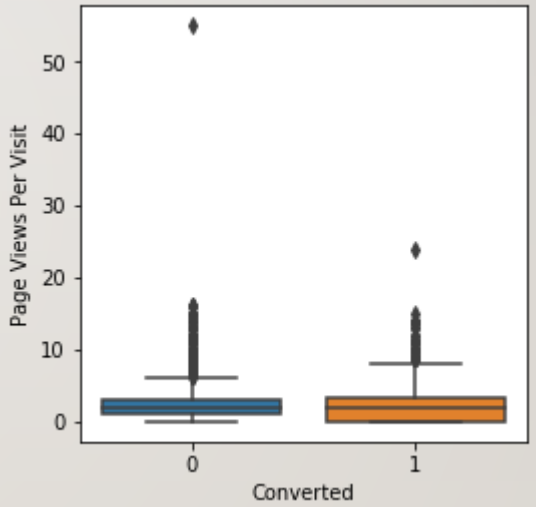
Final Model Analysis and performance on Test Data

OUTLIERS

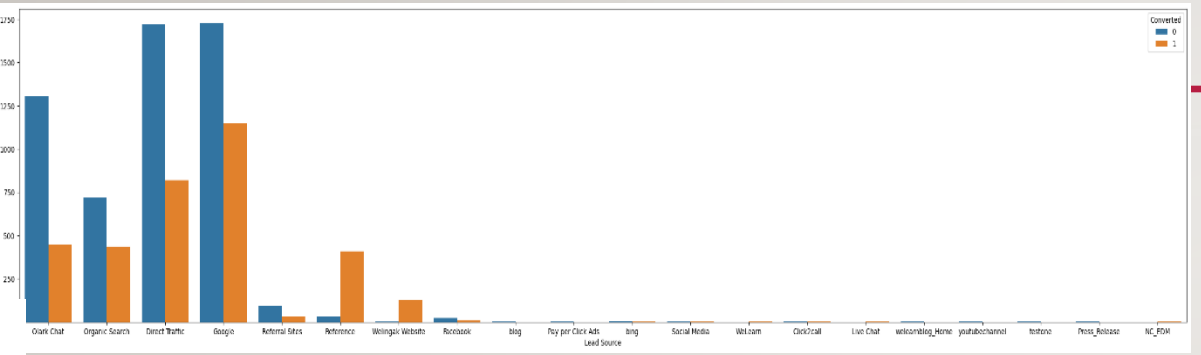
- OUTLIERS IN NUMERICAL COLMUNS



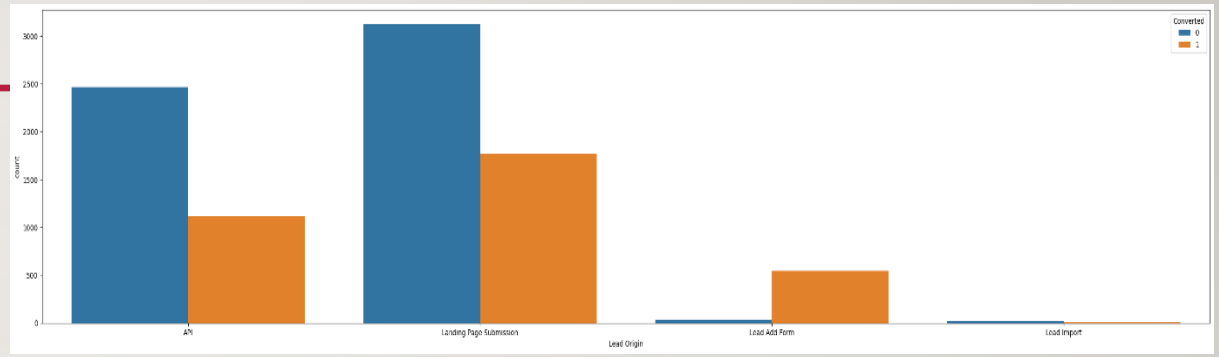




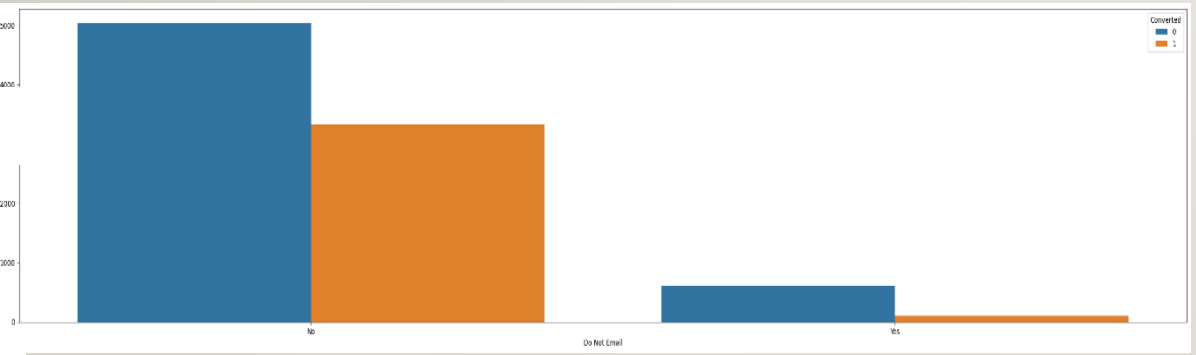
COUNT PLOTS OF CATAGORICAL DATA



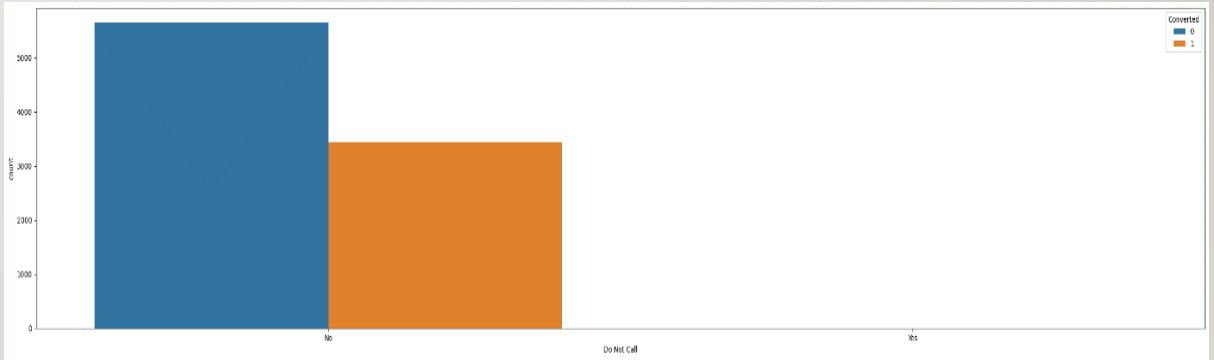
LEAD SCORE



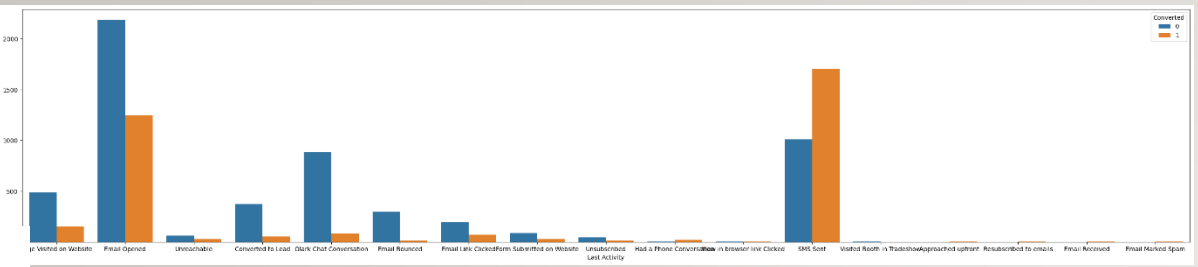
LEAD ORIGIN



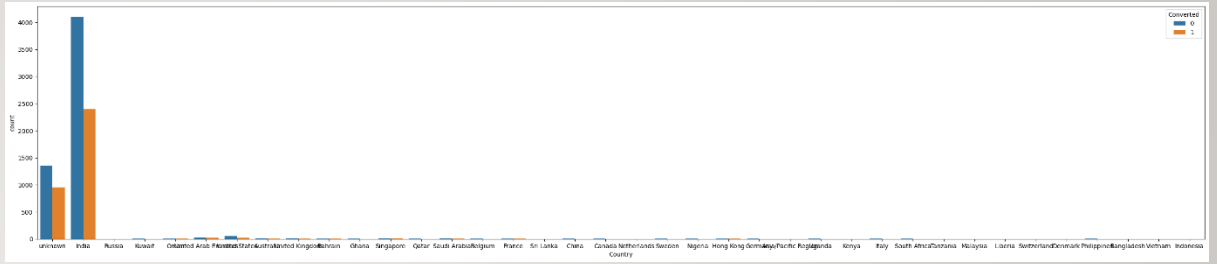
DO NOT EMAIL



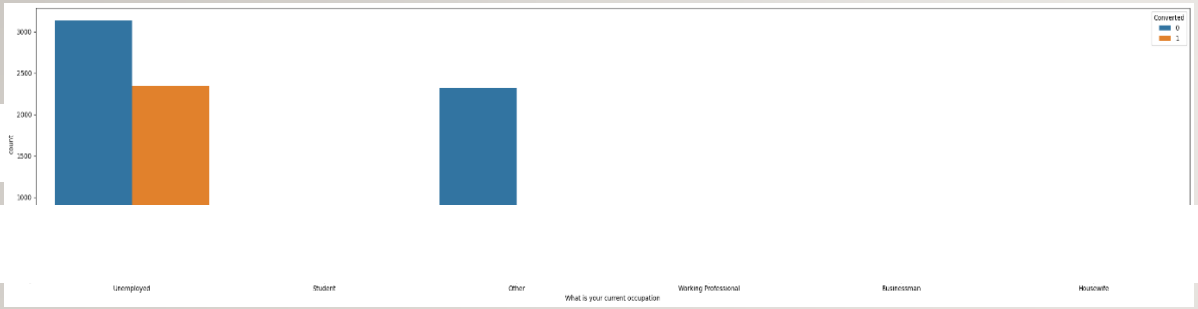
DO NOT CALL



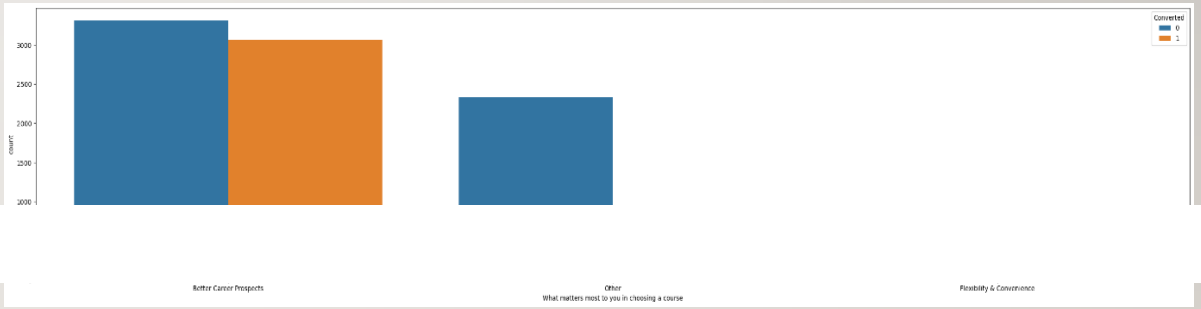
LAST ACTIVITY



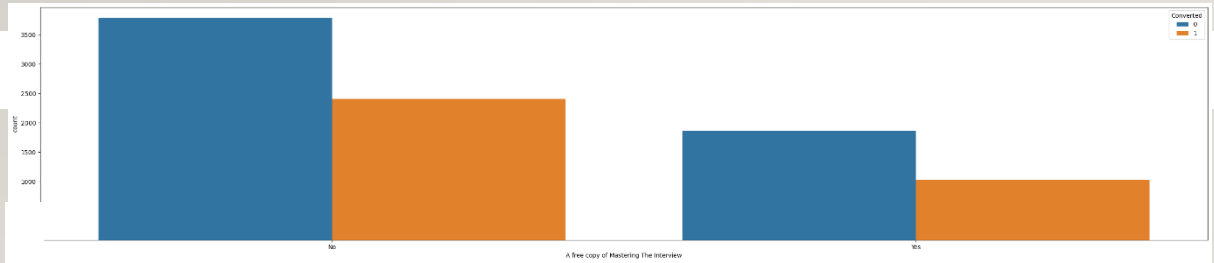
COUNTRY



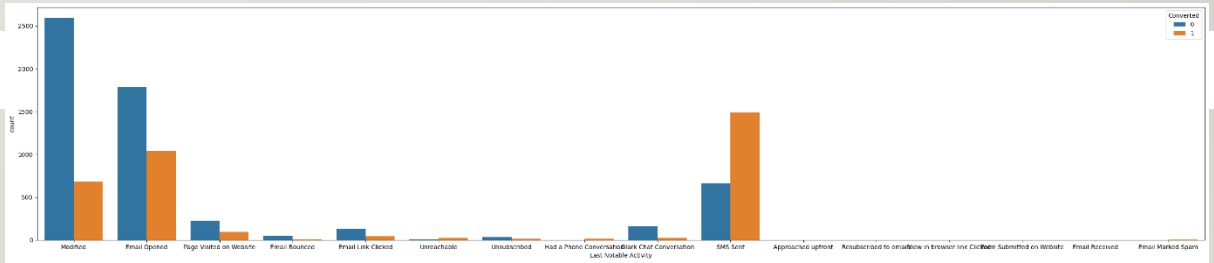
CURRENT OCCUPATION



WHAT MATTERS TO CHOOSE COURSE



FREE COPY OF MASTERING INTERVIEW

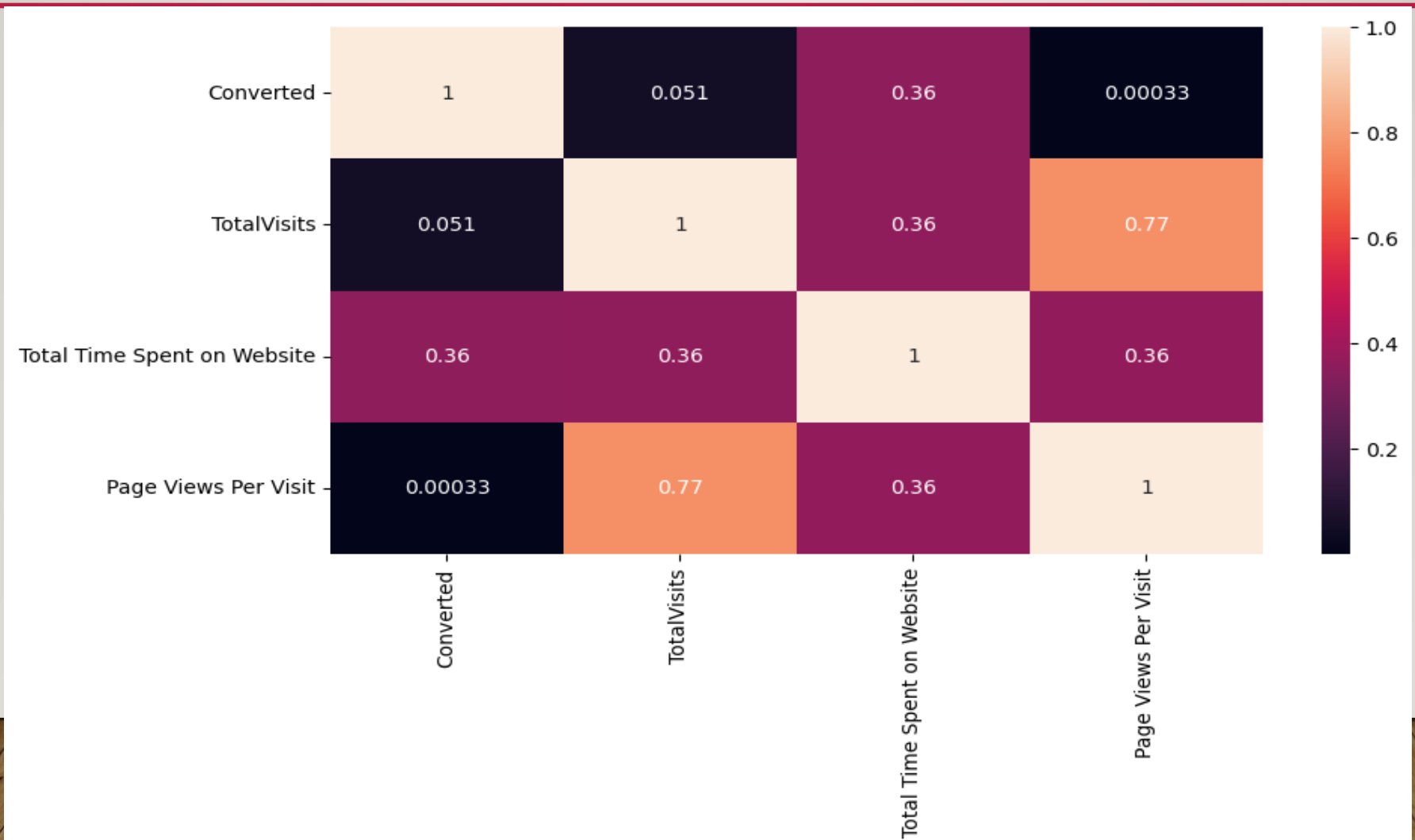


LAST NOTABLE ACTIVITY

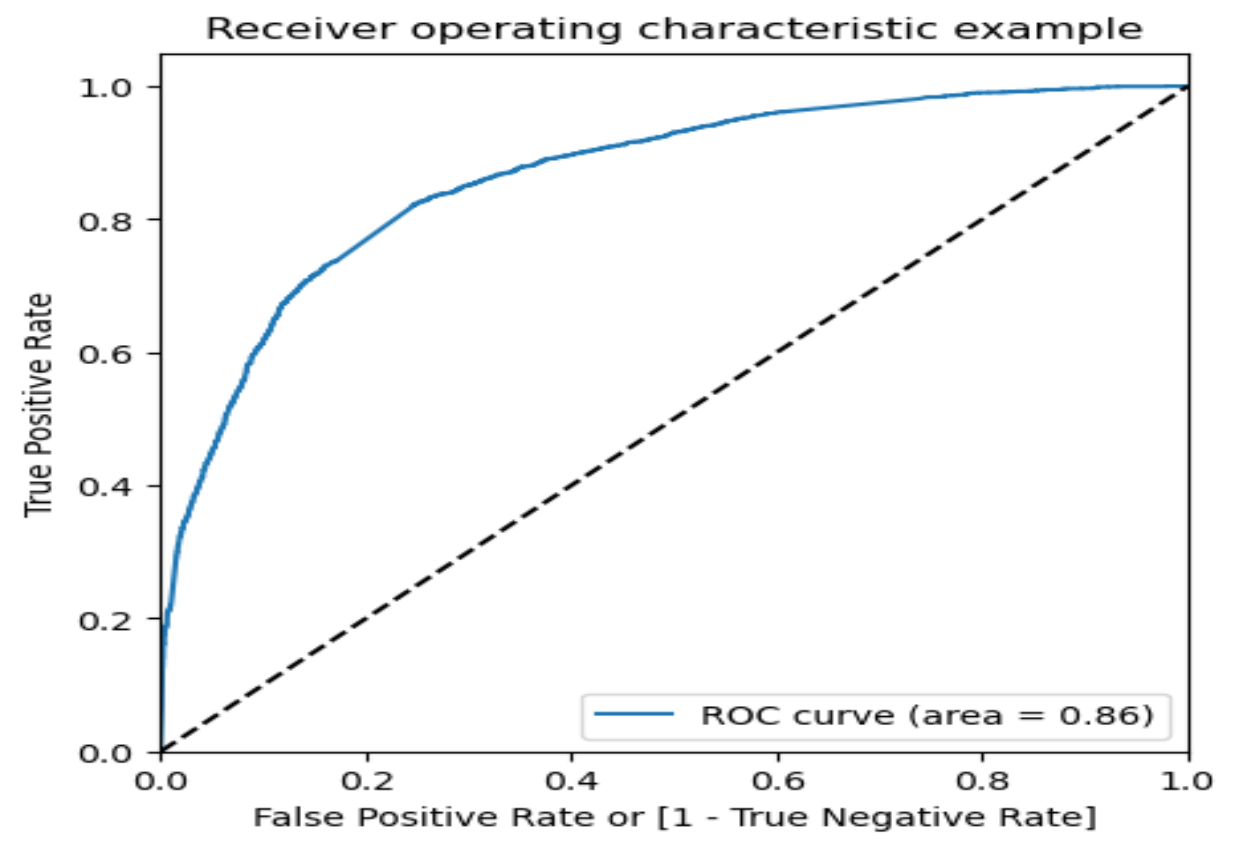
OBSERVATIONS

- 'Lead Add Form' have high conversion rate.
- Max leads are from 'Direct Traffic' and 'Google'
- 'Reference' and 'Welingak website' have high conversion rate
- 'Email Opened' Leads are max and 'SMS Sent' have high conversion rate
- Max Leads are from 'India' but many other seems to be 'unknown'
- Max Leads are 'Unemployed' and some of them haven't mentioned that is 'Other'
- Working Professional' have high conversion rate
- We are dropping other variables due to data imbalance and insignificance

CORRELATIONS



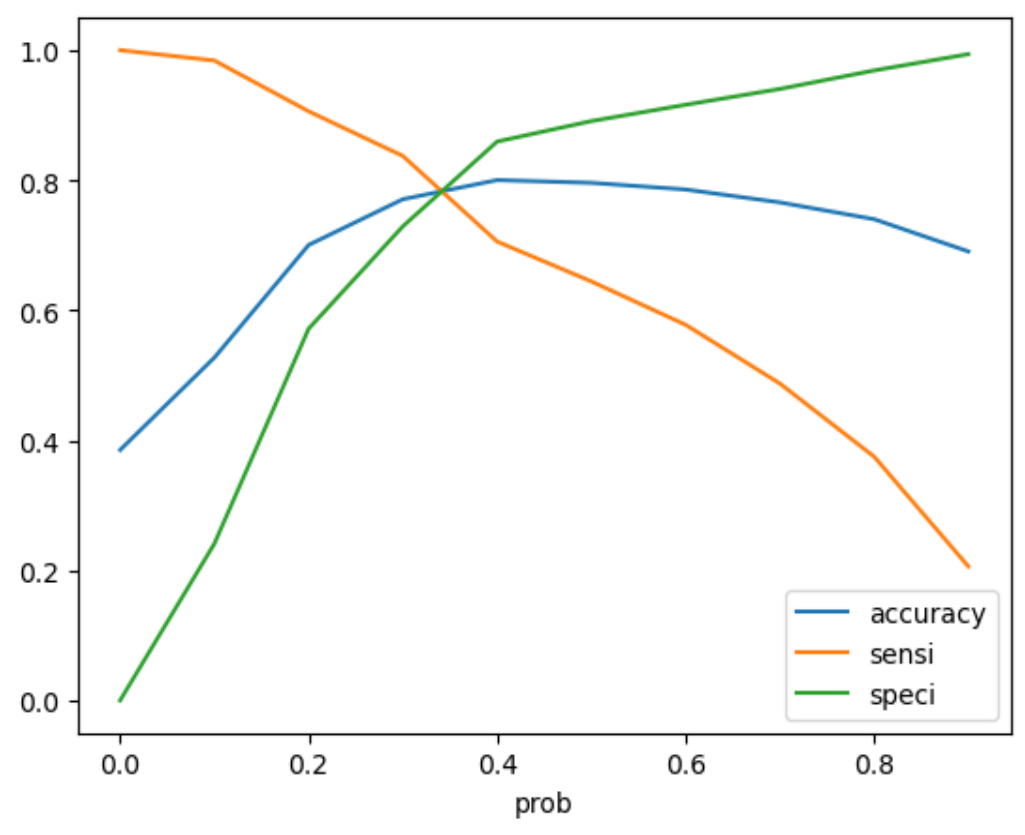
ROC



The Receiver Operating Characteristic (ROC) curve is a graphical representation of the performance of a classification model across various thresholds. The Area Under the ROC Curve (AUC) is a metric that quantifies the overall performance of the model. An AUC value of 0.86 indicates that the model has a high discriminatory power, with a strong ability to distinguish between positive and negative classes.

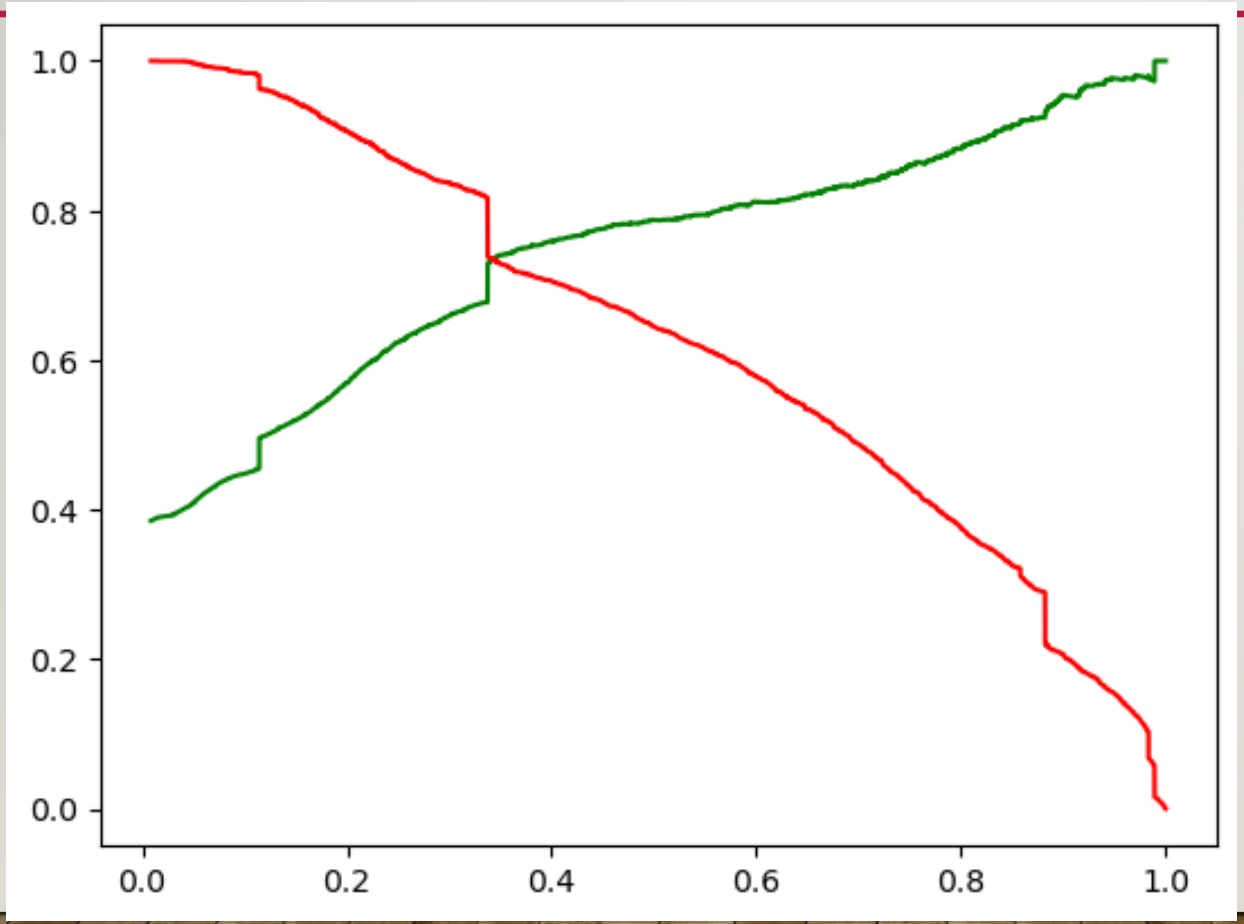
This performance level suggests that the model built by X Education for identifying hot leads is quite effective, with a high likelihood of accurately classifying leads into their respective categories.

CUTOFF POINT



	prob	accuracy	sensi	speci
0.0	0.0	0.385136	1.000000	0.000000
0.1	0.1	0.527161	0.984056	0.240973
0.2	0.2	0.700362	0.905969	0.571575
0.3	0.3	0.770745	0.837285	0.729065
0.4	0.4	0.800189	0.705642	0.859411
0.5	0.5	0.795938	0.644317	0.890909
0.6	0.6	0.785703	0.577678	0.916005
0.7	0.7	0.765864	0.487326	0.940333
0.8	0.8	0.740041	0.374898	0.968758
0.9	0.9	0.690442	0.206051	0.993854

PRECISION AND RECALL TRADEOFF



MODEL ANALYSIS

Performance of our Final Model

Overall accuracy on Test set: 77%

Sensitivity of our logistic regression model: 83%

Specificity of our logistic regression model: 73%

INFERENCES FROM MODEL

Business Insights Derived from our Model

Top 3 variables in model, that contribute towards lead conversion are:

- Total Time Spent on Website 4.645322
- Lead Origin_Lead Add Form 4.107177
- What is your current occupation_Working Professional 2.471612

RECOMMENDATIONS

- Focus on Website Engagement: Invest in strategies to enhance the user experience on the website and encourage visitors to explore more pages. Engaging content, clear navigation, and personalized recommendations can help increase website engagement.
- Targeted Marketing Campaigns: Tailor marketing campaigns based on lead source and occupation preferences. Allocate resources to channels with high conversion rates, such as reference sites and specific professions like working professionals.
- Lead Follow-up: Prioritize leads with higher lead scores for follow-up actions. Implement lead nurturing strategies to maintain engagement with leads over time, providing relevant information and addressing their concerns.
- Continuous Model Monitoring: Regularly monitor the performance of the lead scoring model and update it as necessary. Incorporate new data and refine the model to improve its predictive accuracy over time.
- Collaboration between Sales and Marketing: Foster collaboration between sales and marketing teams to ensure seamless lead management processes. Align on lead definitions, scoring criteria, and communication strategies to maximize conversion opportunities.
- Customer Feedback: Gather feedback from converted leads to understand their journey and identify areas for improvement. Use insights to refine marketing strategies, enhance product offerings, and optimize the overall customer experience

CONCLUSION

- Lead Scoring Model: The lead scoring model developed in this analysis provides a systematic approach to predict the conversion probability of leads. By considering various features such as lead source, occupation, and website engagement metrics, the model identifies potential leads more likely to convert.
- Model Evaluation: The model achieved an accuracy score of approximately 82% on the test dataset, indicating its effectiveness in predicting conversions. Additionally, the sensitivity and specificity scores indicate a balanced performance in identifying true positives and true negatives.
- Key Variables: The analysis identified several key variables that significantly contribute to conversion probability. These include total time spent on the website, page views per visit, and certain lead sources and occupations. Leads who spend more time on the website and visit more pages per visit are more likely to convert.

CONCLUSION

- Lead Scoring Model: The lead scoring model developed in this analysis provides a systematic approach to predict the conversion probability of leads. By considering various features such as lead source, occupation, and website engagement metrics, the model identifies potential leads more likely to convert.
- Model Evaluation: The model achieved an accuracy score of approximately 82% on the test dataset, indicating its effectiveness in predicting conversions. Additionally, the sensitivity and specificity scores indicate a balanced performance in identifying true positives and true negatives.
- Key Variables: The analysis identified several key variables that significantly contribute to conversion probability. These include total time spent on the website, page views per visit, and certain lead sources and occupations. Leads who spend more time on the website and visit more pages per visit are more likely to convert.

THANK YOU