

LEAD SCORING CASE STUDY

BY: Rohan Pandey


PROBLEM STATEMENT

- X Education in the Ed-Tech Industry sells online courses to industry professionals.
- It's lead conversion is very poor.
- The company wants to assign Lead Score to each lead using Machine Learning Model.

Assumptions

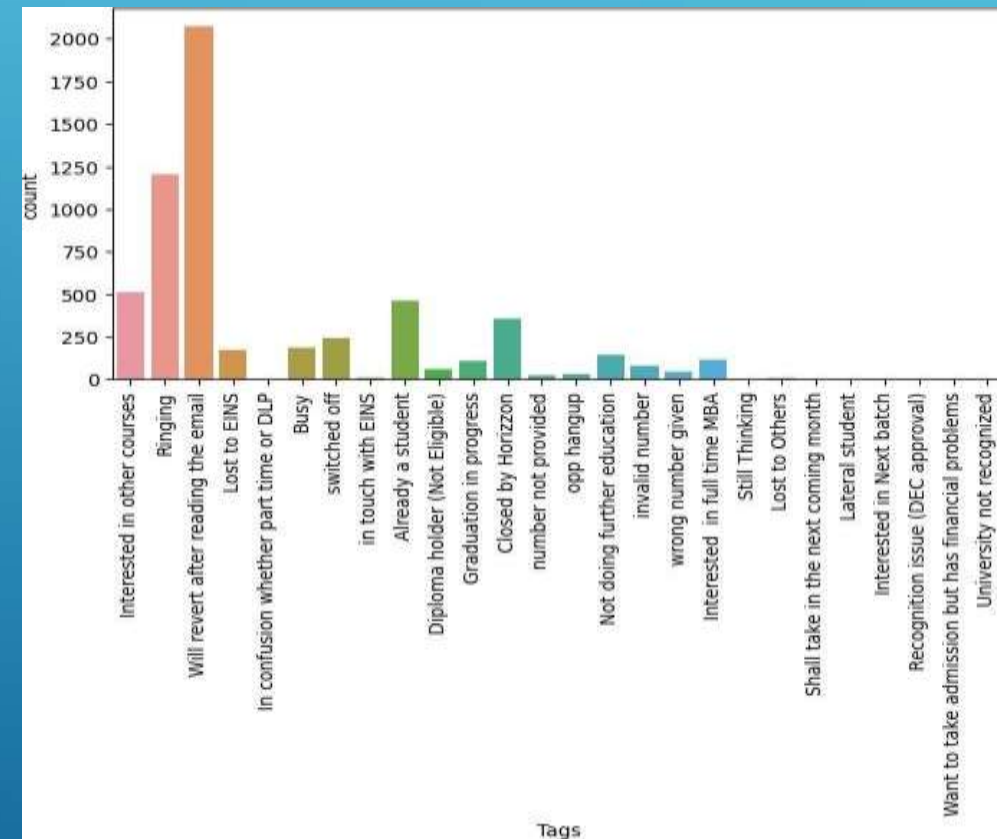
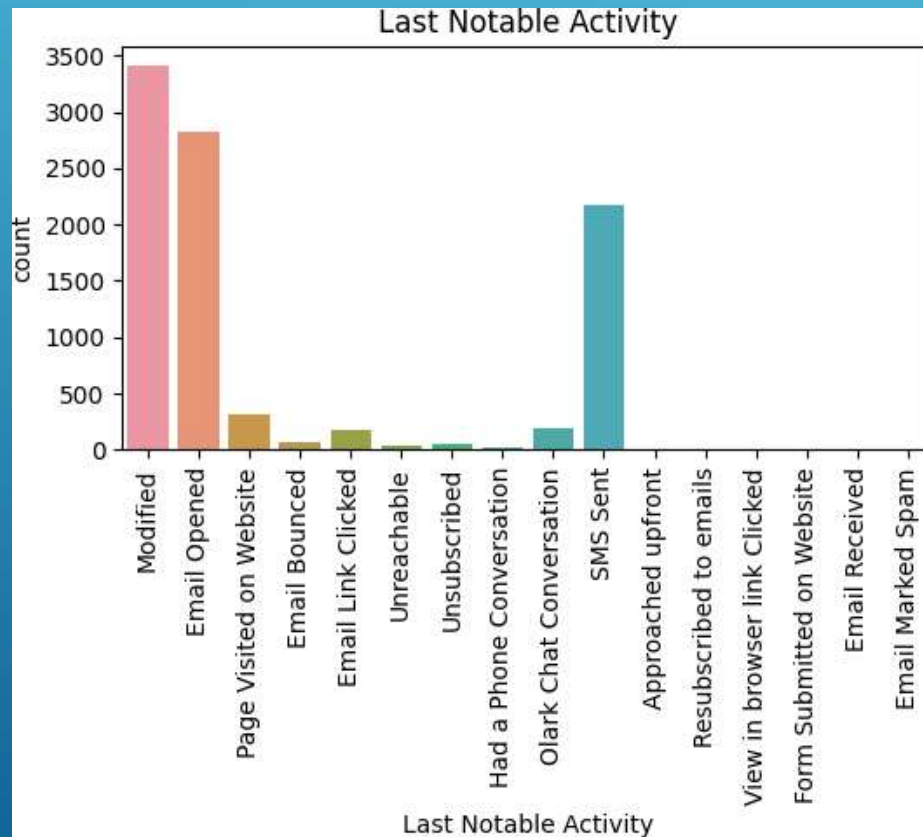
- Unique value variables like 'Prospect ID' and Single value variables are dropped as they do not provide any significant information.
- Variables with high missing values and data imbalances are also not considered.

APPROACH FOLLOWED

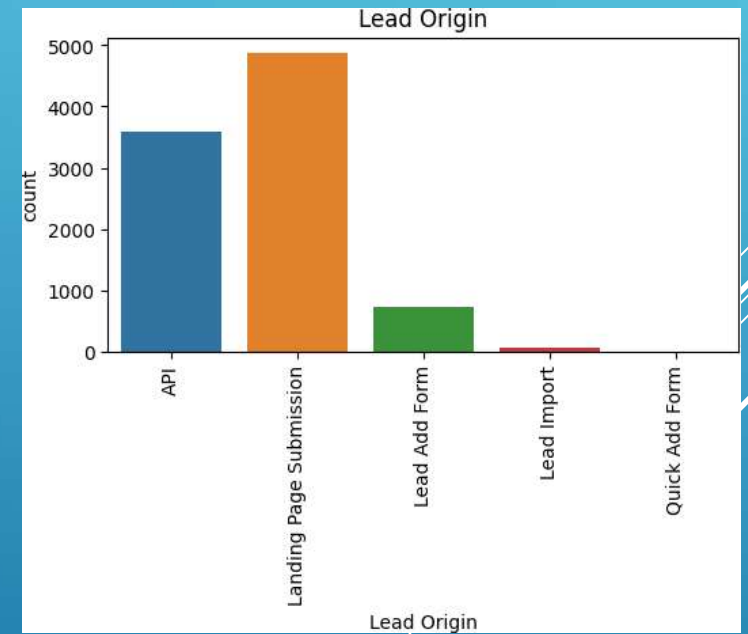
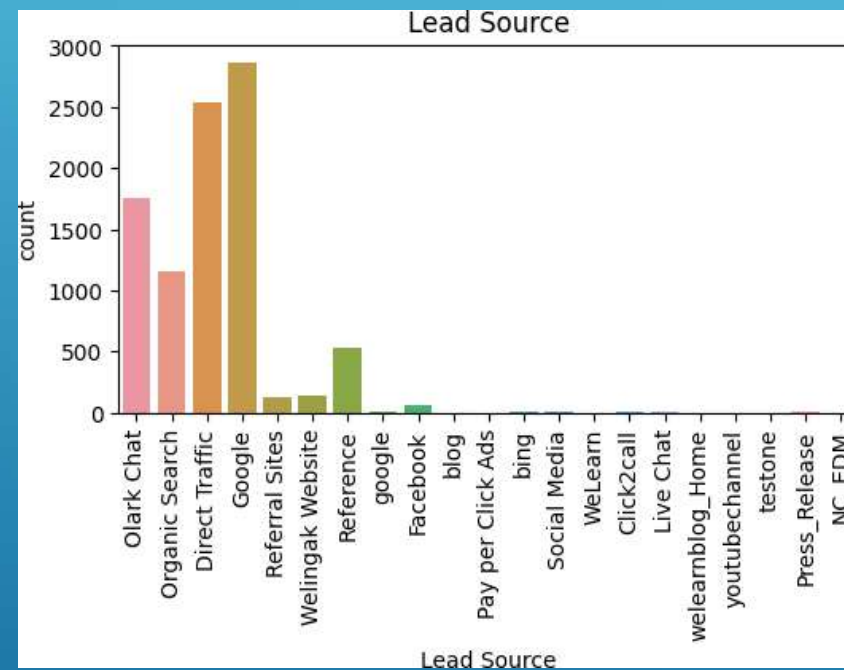
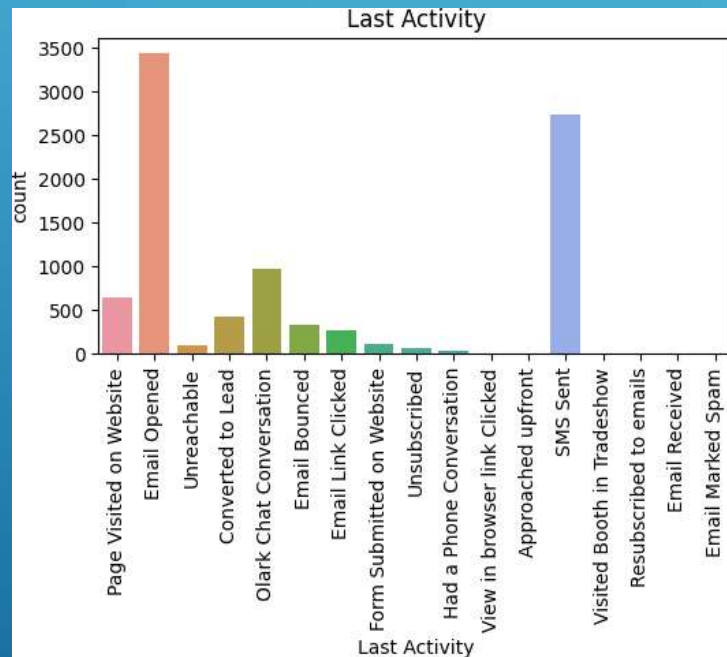
1. Data Cleaning and Data Manipulation
 2. Exploratory Data Analysis
 3. Data Pre-processing
 4. Model Building
 - Logistic Regression Model
 5. Model Evaluation
 6. Predictions
 7. Conclusions and Recommendations
- 
- Several white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

Notable EDA Conclusions

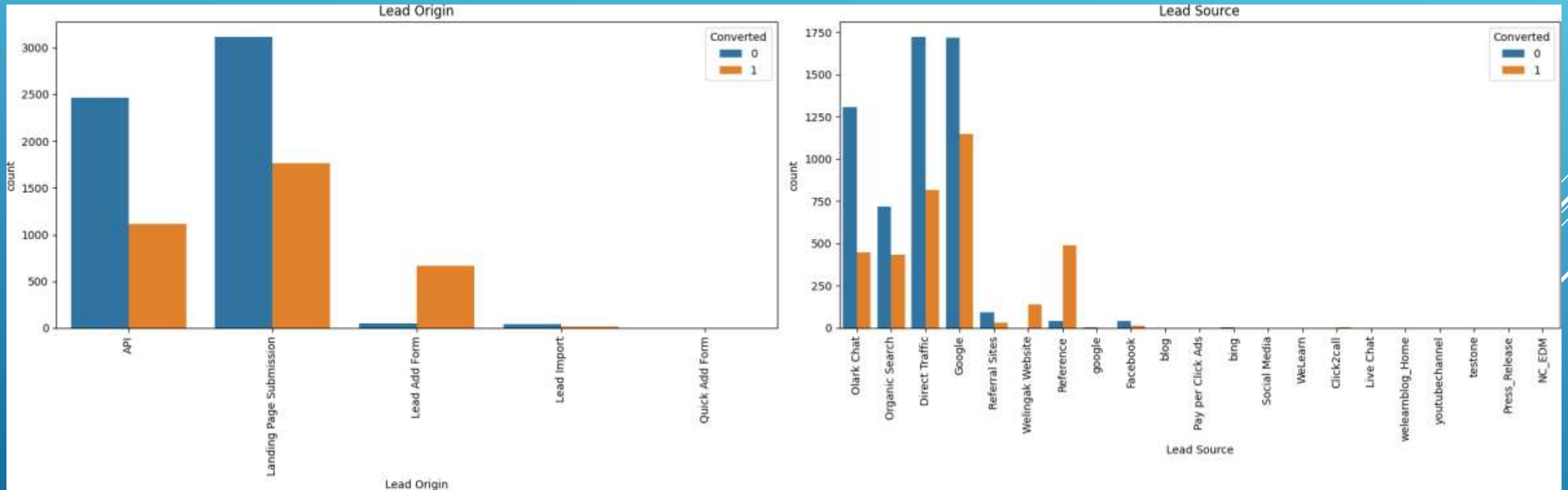
The following variables seem to give significant information about the dataset



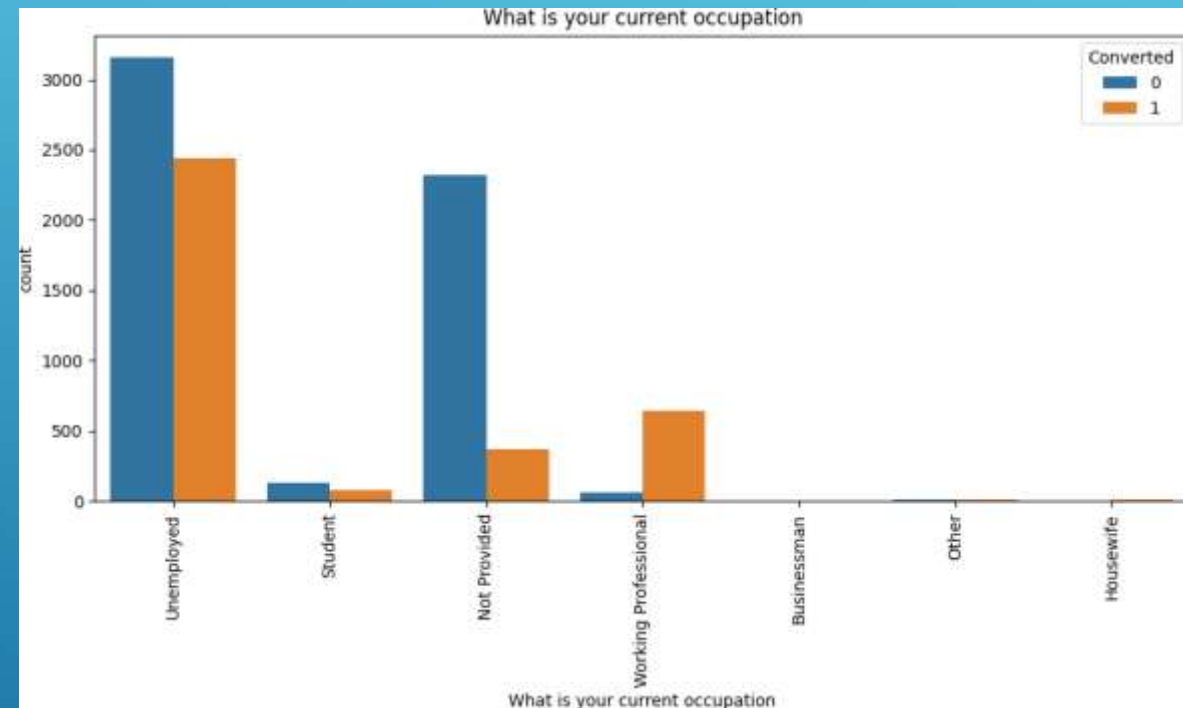
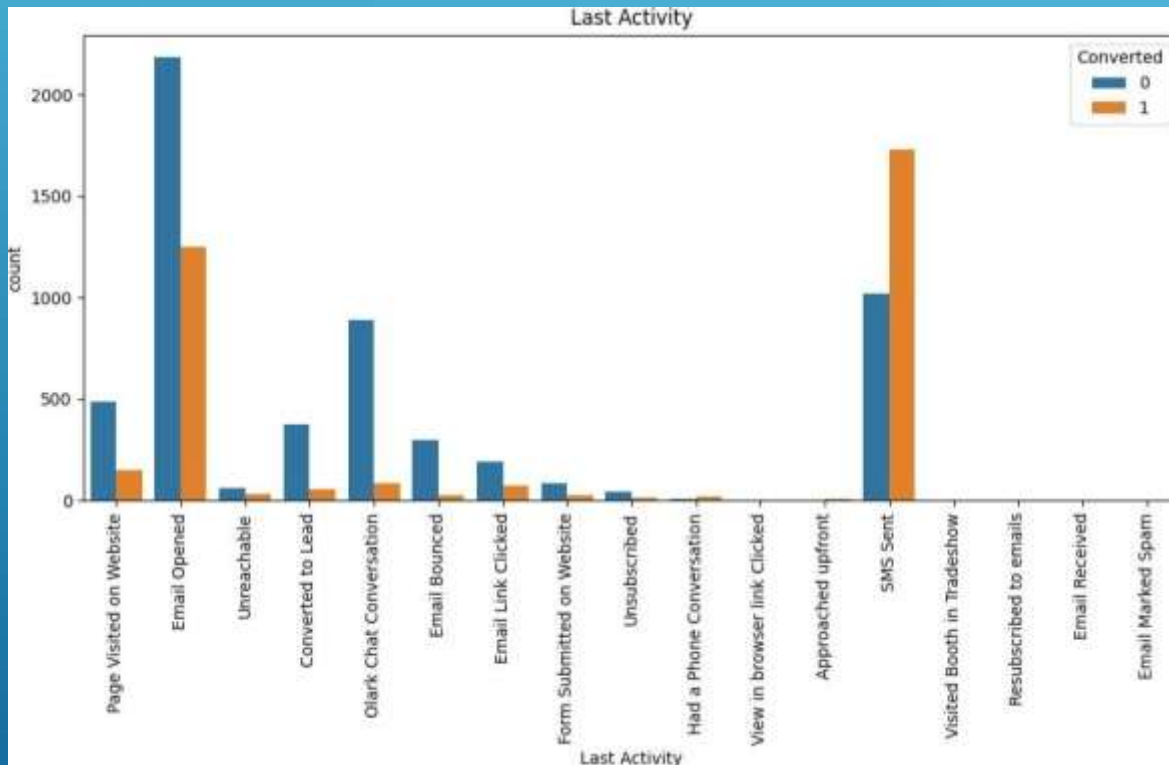
NOTABLE EDA CONCLUSIONS (CONT.)



NOTABLE EDA CONCLUSIONS (CONT.)



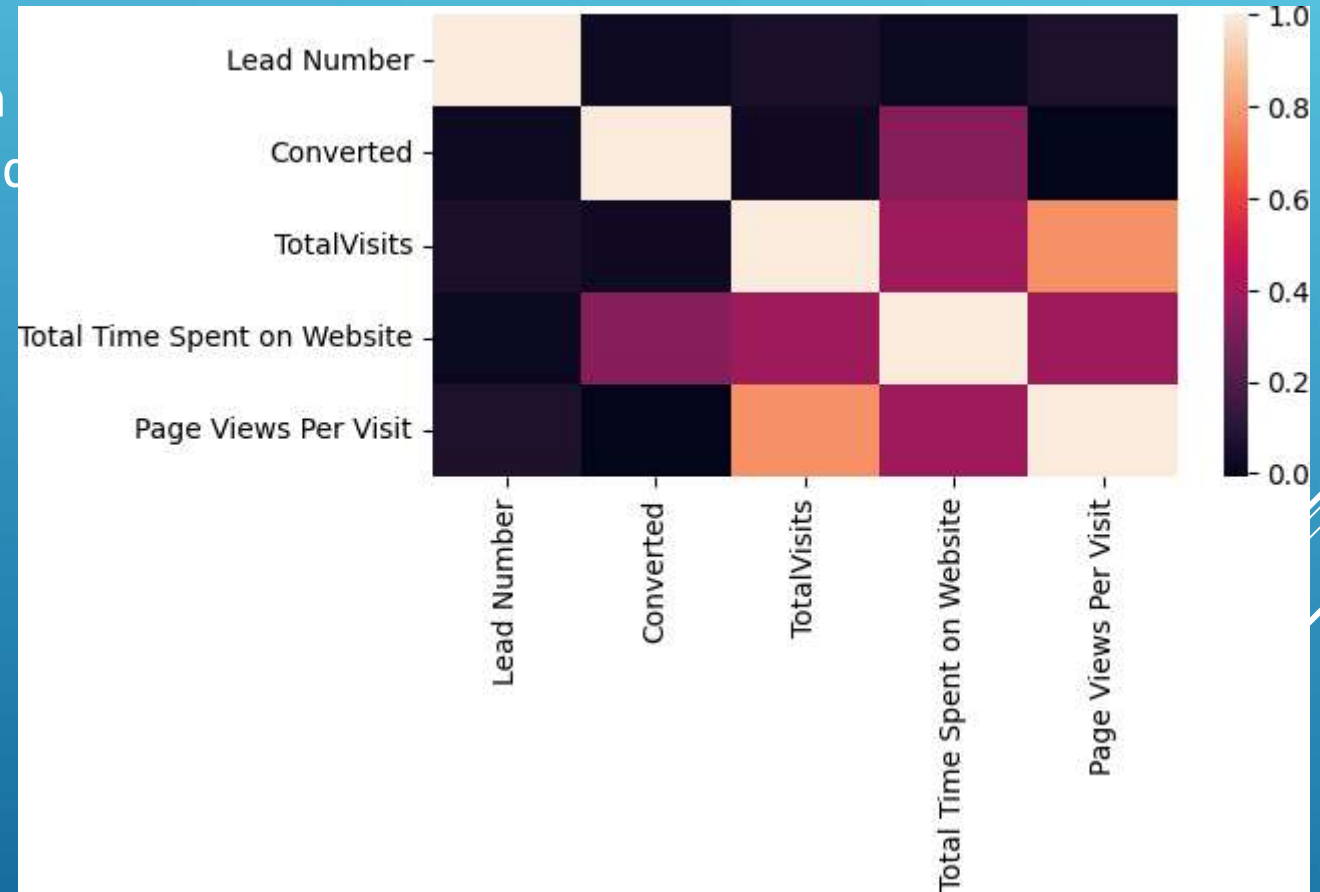
NOTABLE EDA CONCLUSIONS (CONT.)



NOTABLE EDA CONCLUSIONS (CONT.)

This correlation matrix displays high correlation between 'Converted' and

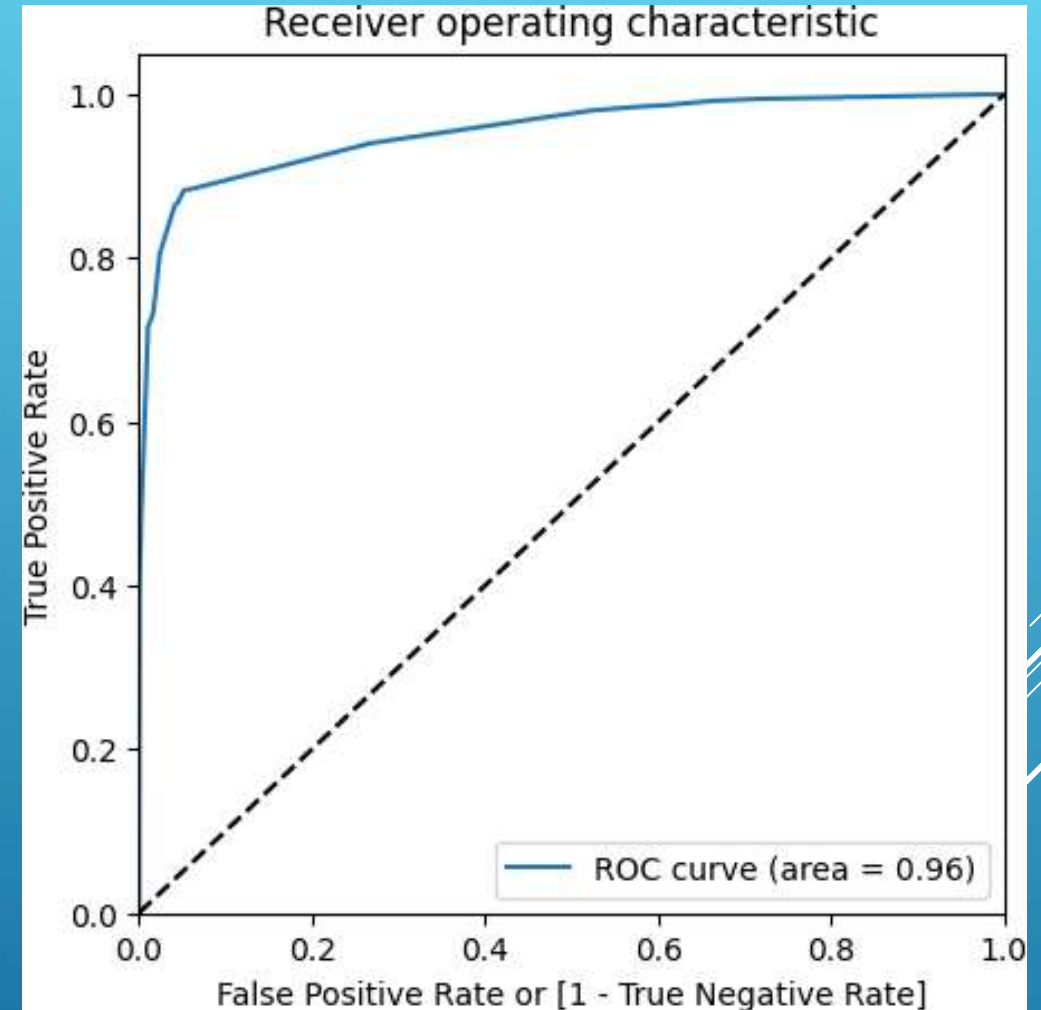
- Total visits
- Page views per visit
- Time spend on website



RESULTS

ROC Curve

- 96% of the area is under ROC curve.
- Classification Probability of lead conversion by the model is very high.



RESULTS

CONFUSION

Actual/Predicted	Not Converted	Converted
Not Converted	3499	204
Converted	259	1969

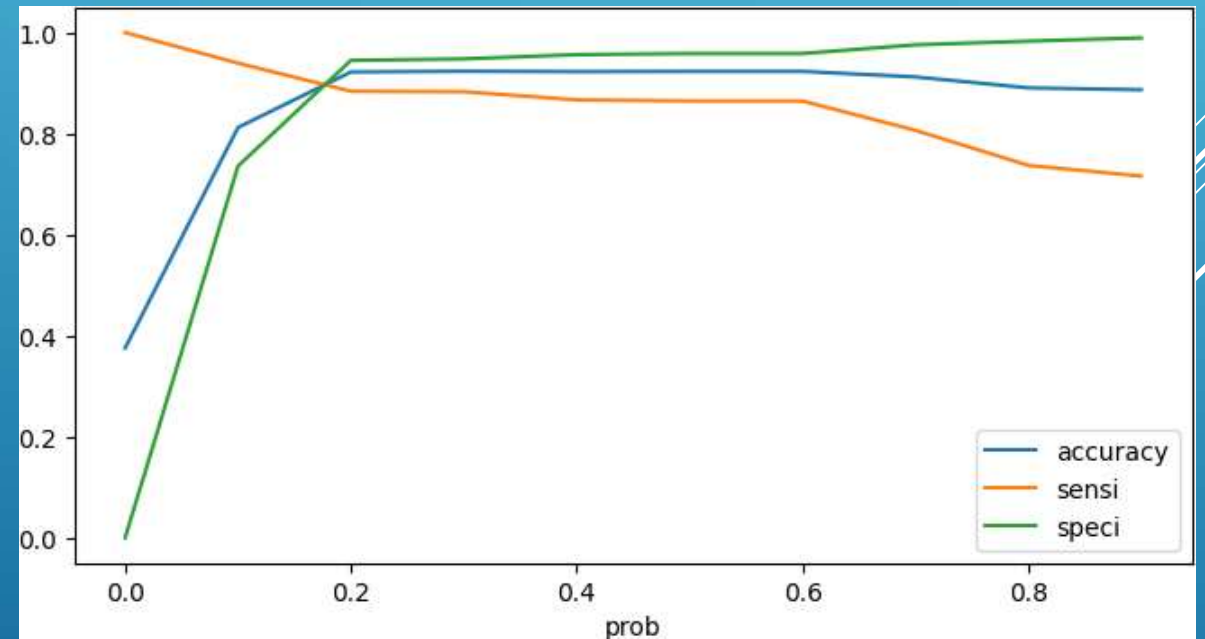
Optimal Probability Cut-off

With 0.2 cut-off, the model has:

Accuracy – 92%

Sensitivity – 88%

Specificity – 94%



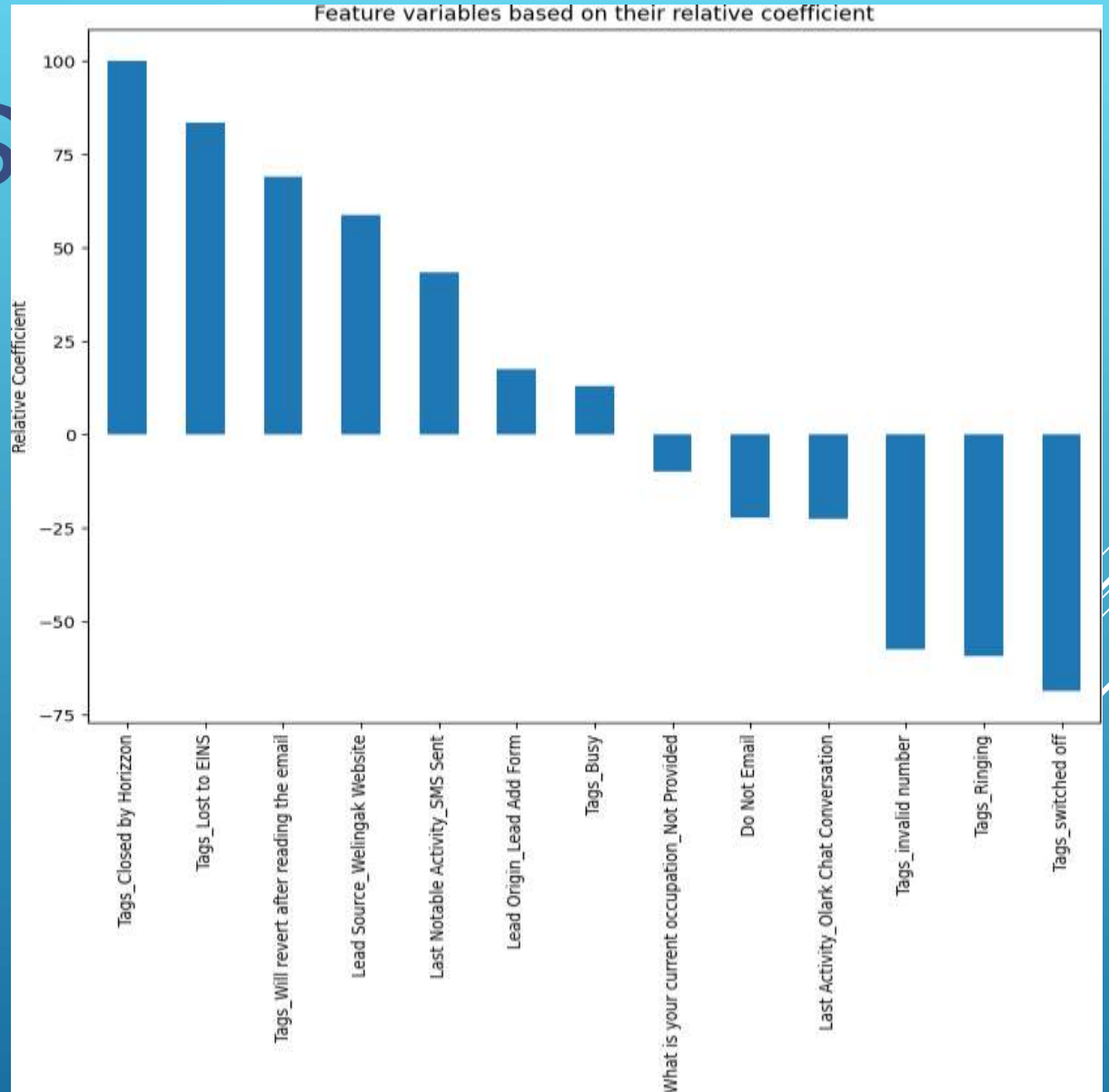
IMPORTANT FEATURES

Top 3 variables that have **high lead conversion probability**

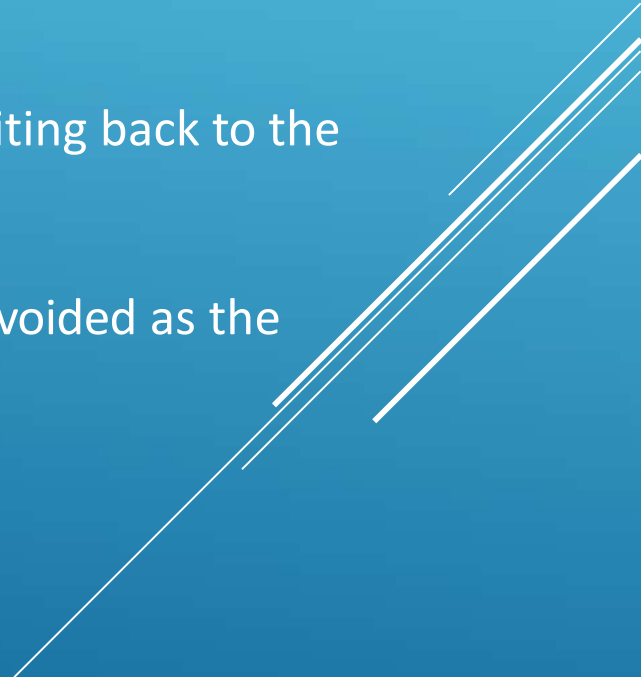
- Tags – Closed by Horizon, Lost to EINS, Will revert after reading the email
- Lead Source – Welingak Website
- Last Notable Activity – SMS Sent

Top 3 variables that **need improvement** in converting quality lead

- Tags – Invalid Number
- Tags – Ringing
- Tags – Switched off



RECOMMENDATIONS

- **Follow ups** through calls and emails with high conversion probability leads is suggested.
 - Focus more on customers who spend a lot of time on the company's website as their conversion rate is high as per EDA.
 - Providing special offers to customers who are highly interested and are seen visiting back to the website.
 - Leads who have Tags such as 'Ringing', 'Switched Off', 'Invalid Number' can be avoided as the probability of them converting is very low.
- 
- A series of white diagonal lines of varying lengths and thicknesses are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

THANK YOU

