

# REPORT

## Exploratory Data Analysis Summary

1. **Initial Cleaning:** Removed columns containing unique IDs and Lead numbers as they were not relevant to the analysis.
2. **Missing Data Handling:**
  - Replaced instances of "Select" with null values, indicating no selection was made.
  - Dropped columns with over 40% missing data.
  - Imputed missing values in columns with less than 10% missingness using mean, median, or mode, depending on the data type.
3. **Feature Refinement:**
  - Examined column distributions and removed those with excessive skewness (asymmetry).
  - Combined relevant categories for simplified analysis.
4. **Outlier Treatment & Dummy Variables:** Managed outliers in numerical columns and created dummy variables for categorical data.
5. **Visualization:** Created various plots to visualize data distributions and relationships.

## 2. Data Preparation :

- We transformed binary columns such as 'Do Not Email' & 'A free copy of Mastering The Interview' to '0' & '1' from 'No' & 'Yes'
- Then we created dummy variables

## 3. Test-train Split :

- We splitted the data into Test-train in the proportion of 30% and 70%.

## 4. Model Building & Evaluation

- Finally we did precision-recall tradeoff which gives us accuracy of 78%, precision and recall of 76% and 66% on training set.

## CONCLUSION

The Model seems to predict the Conversion Rate very well and we should be able to give the Company confidence in making good calls based on this model.

The variables that mattered the most in the potential buyers are :

# Total Time Spent on Website

# Page Views Per Visit

# Do Not Email

# What is your current occupation\_Working Professional

# Lead Origin\_Landing Page Submission

# Specialization\_Industry Specializations

# A free copy of Mastering The Interview

# TotalVisits