



# Lead Score Case Study

---

SUBMITTED BY:

SAURABH AMBEKAR

SWAPNIL WADIKAR

SHAHIR MOHAMED T M

# Solution Approach

---

- ☐ Data Inspection
- ☐ Data Cleaning
- ☐ EDA
- ☐ Data Preparation
- ☐ Model Building
- ☐ Prediction & Result Analysis
- ☐ Prediction on Test Set
- ☐ Conclusion

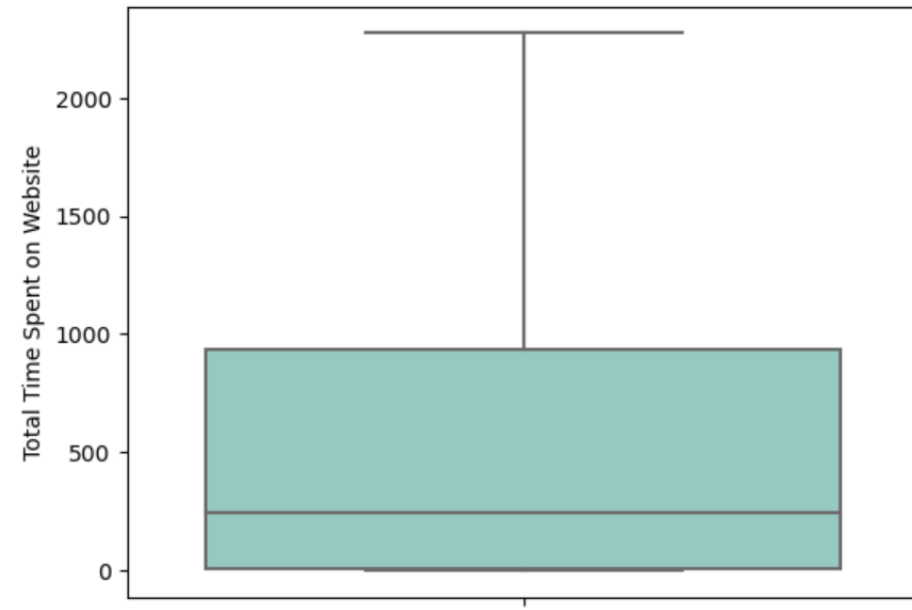
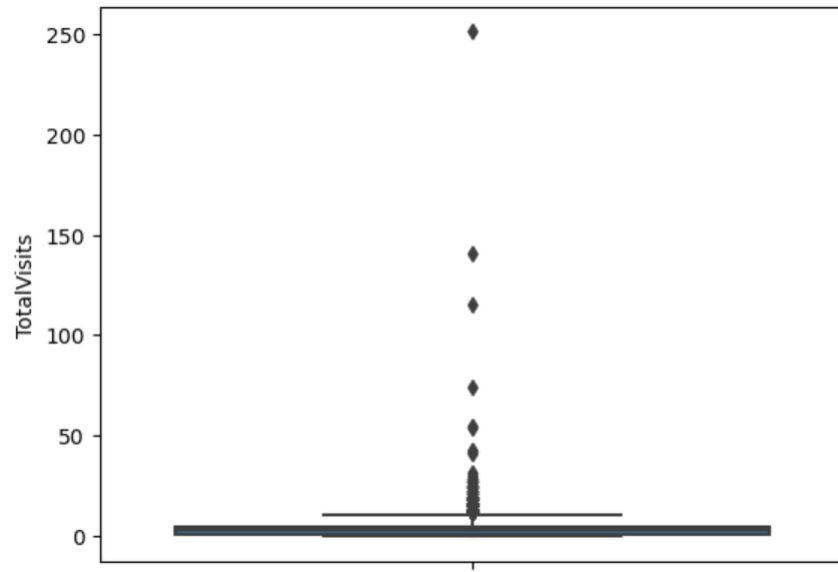
# Data Inspection and Cleaning

---

- ❑ Deleted columns with 35% or more null values
- ❑ Single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply” removed
- ❑ Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis.
- ❑ Dropped columns which had not enough variance, the features are: “Do Not Call”, “What matters most to you in choosing course”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.

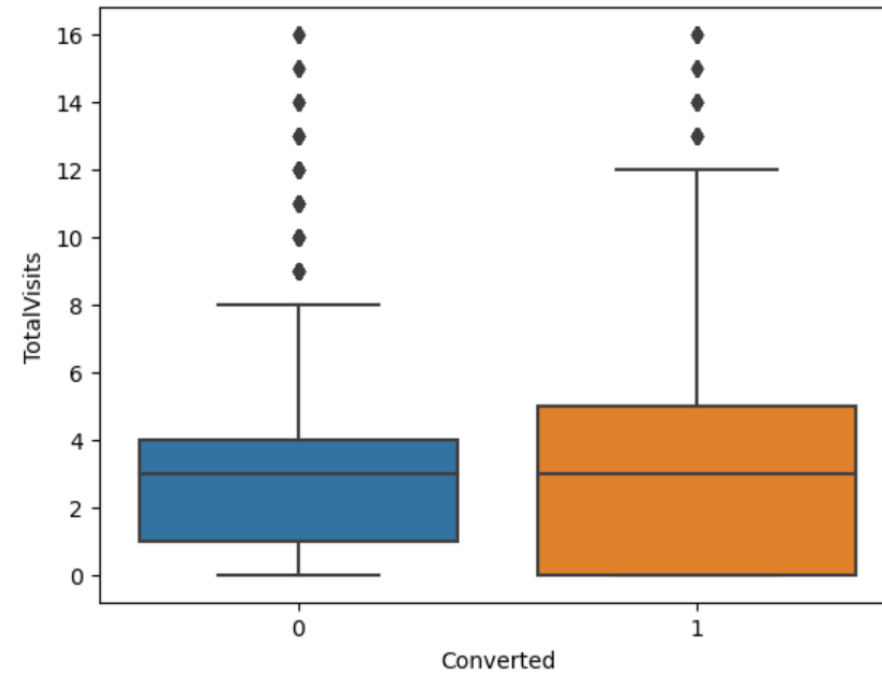
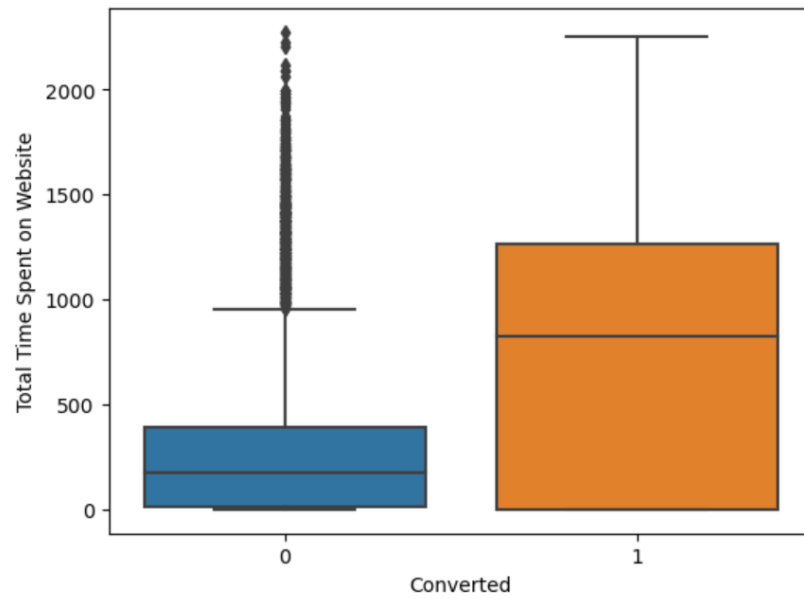
# EDA

## ❑ Box Plot Analysis for Outlier treatment



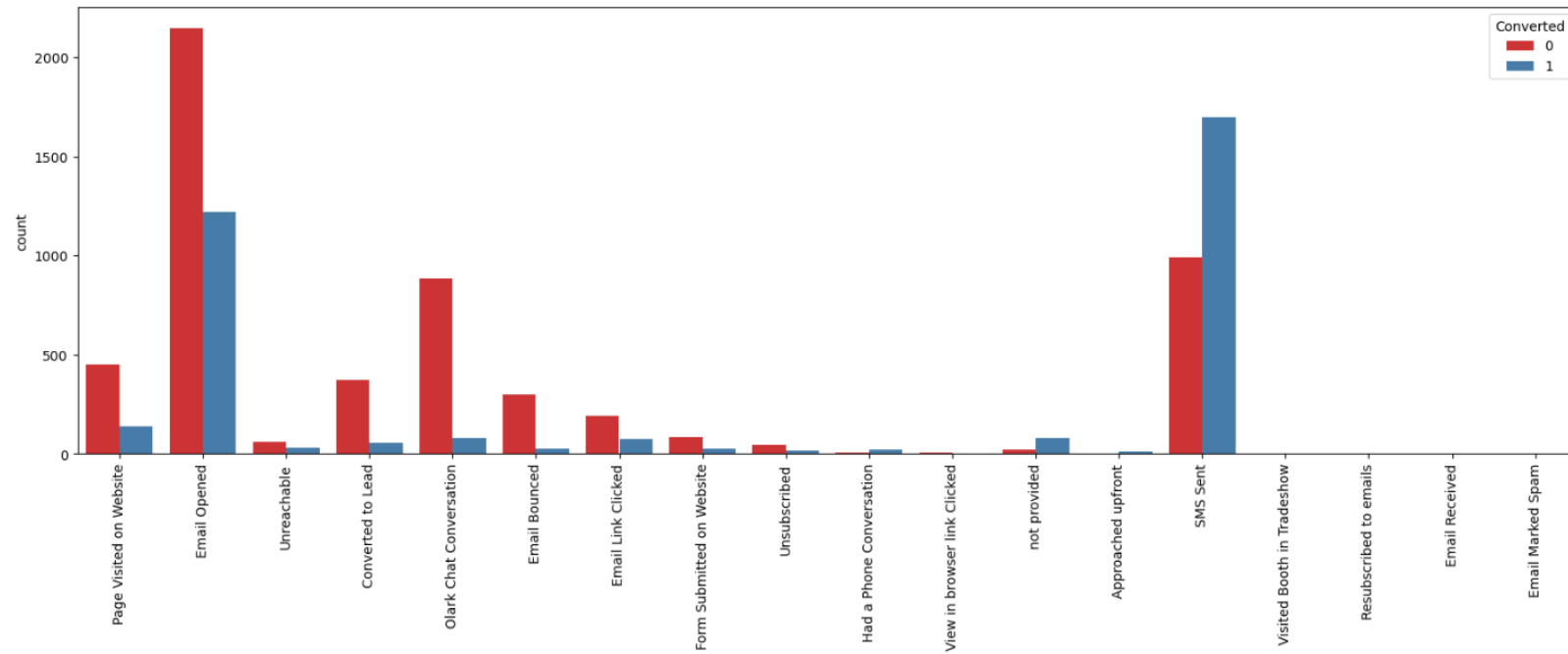
# EDA

❑ Comparing numerical values with target column



# EDA

## □ Categorical variable comparison with the target variable



# Data Preparation

---

- ☐ Converting categorical columns to binary and dummy variables
- ☐ Train Test Split(70:30)
- ☐ Scaling Numerical Variable
- ☐ Feature selection using RFE

# Model Building

---

- ❑ Building model using GLM
- ❑ Checking for p-value to be under 0.05 and VIF under 5
- ❑ Dropping columns and rebuilding the model until the target is achieved.
- ❑ Overall Accuracy achieved 80.43



# Model Building

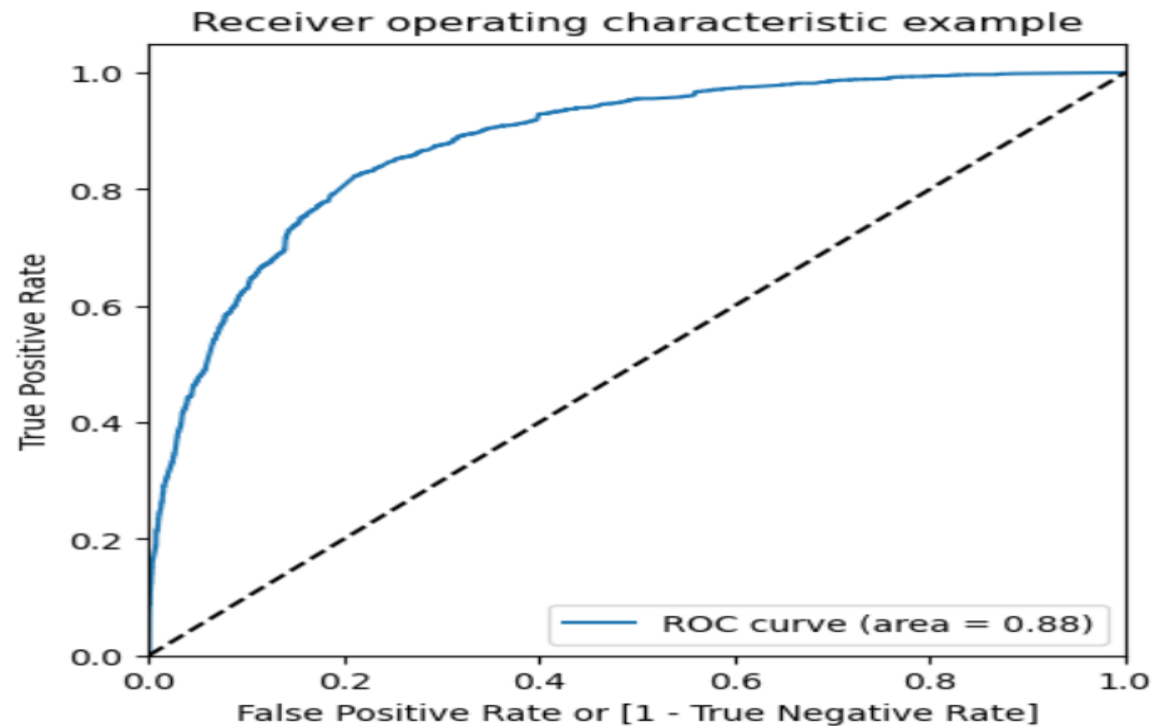
---

- ❑ Building model using GLM
- ❑ Checking for p-value to be under 0.05 and VIF under 5
- ❑ Dropping columns and rebuilding the model until the target is achieved.
- ❑ Overall Accuracy achieved 80.43

# ROC Curve

---

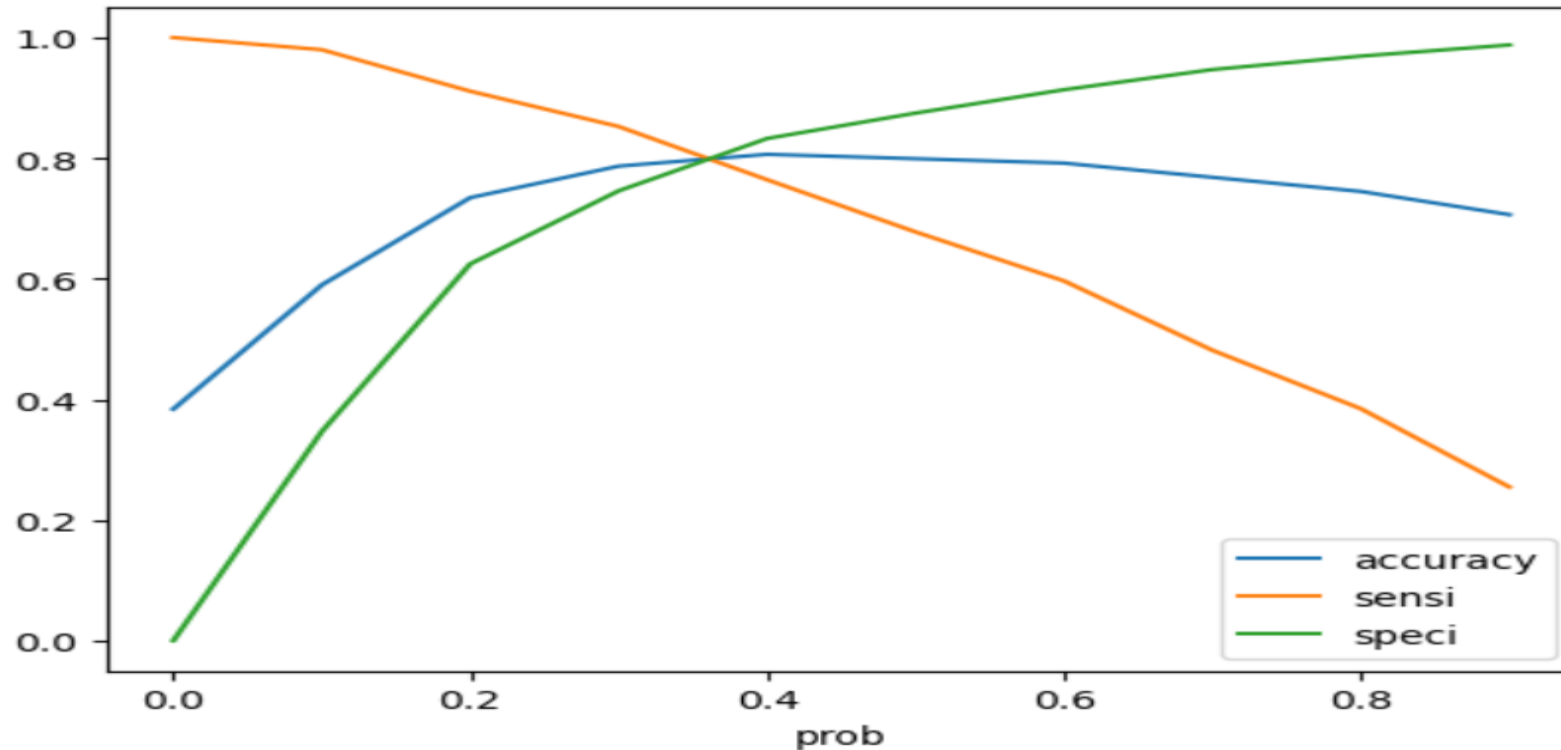
□ Achieved ROC Curve area under 0.88



# Optimal Cut Off Point

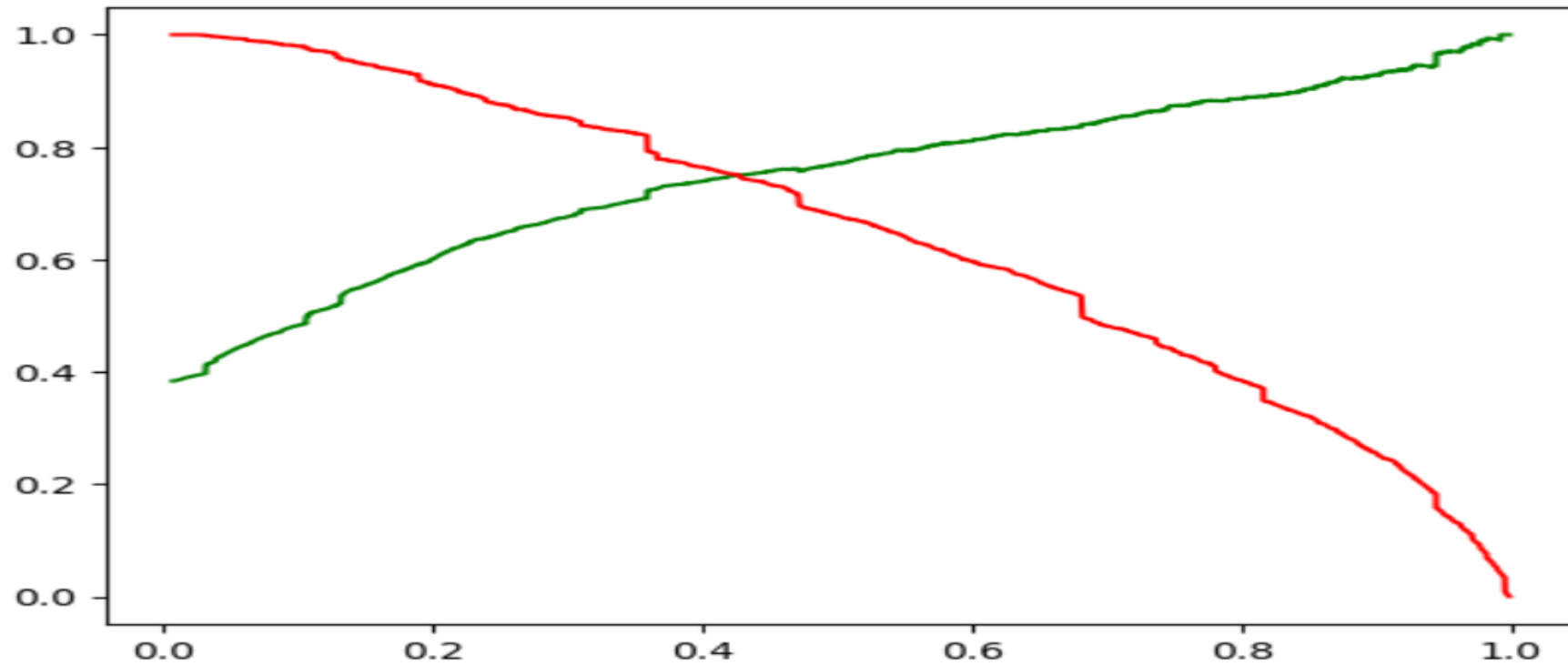
---

❑ Optimal cut-off point was observed to be around 0.37



# Precision Recall Curve

---



# Conclusion

---

- ❑ Accuracy Sensitivity and Specificity for the test set was 79.66, 77.55, 81.0 which is close to the train set
- ❑ Top 3 parameters were
  - ❑ Total Time Spent on Website
  - ❑ Lead Source - Reference
  - ❑ Current occupation - Working Professional
- ❑ Overall model can be considered as good model.

*Thank You!*