LEAD SCORING CASE STUDY

LOGISTIC REGRESSION MODEL

DONE BY :

- NAMRATA NANDI
- BALAJI V
- GAURAV SHARMA

PROBLEM STATEMENT

- X Education sells online courses and gets leads through website visits, referrals, and form submissions.
- They have a typical lead conversion rate of 30%, but wish to identify "Hot Leads" to increase conversion rate. By focusing on potential leads, the sales team can communicate more effectively and convert more leads into paying customers.
- Effective lead nurturing is key in converting potential leads.

BUSINESS GOAL

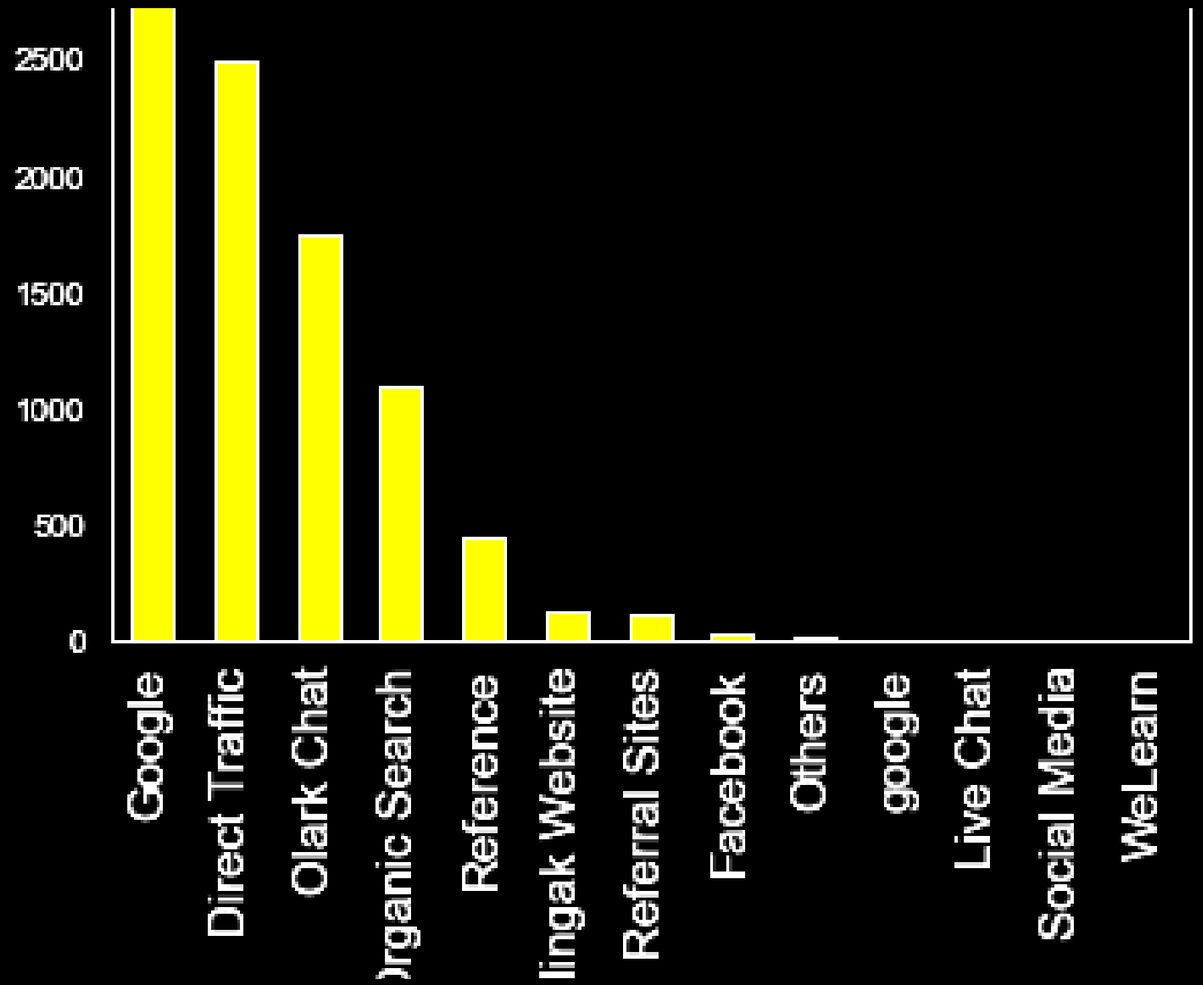
- The CEO of X Education has tasked a consultant with building a lead scoring model to identify potential leads with higher conversion rates.
- The goal is to increase the conversion rate to 80%. The model will assign a lead score to each lead to determine their conversion likelihood.

APPROACH

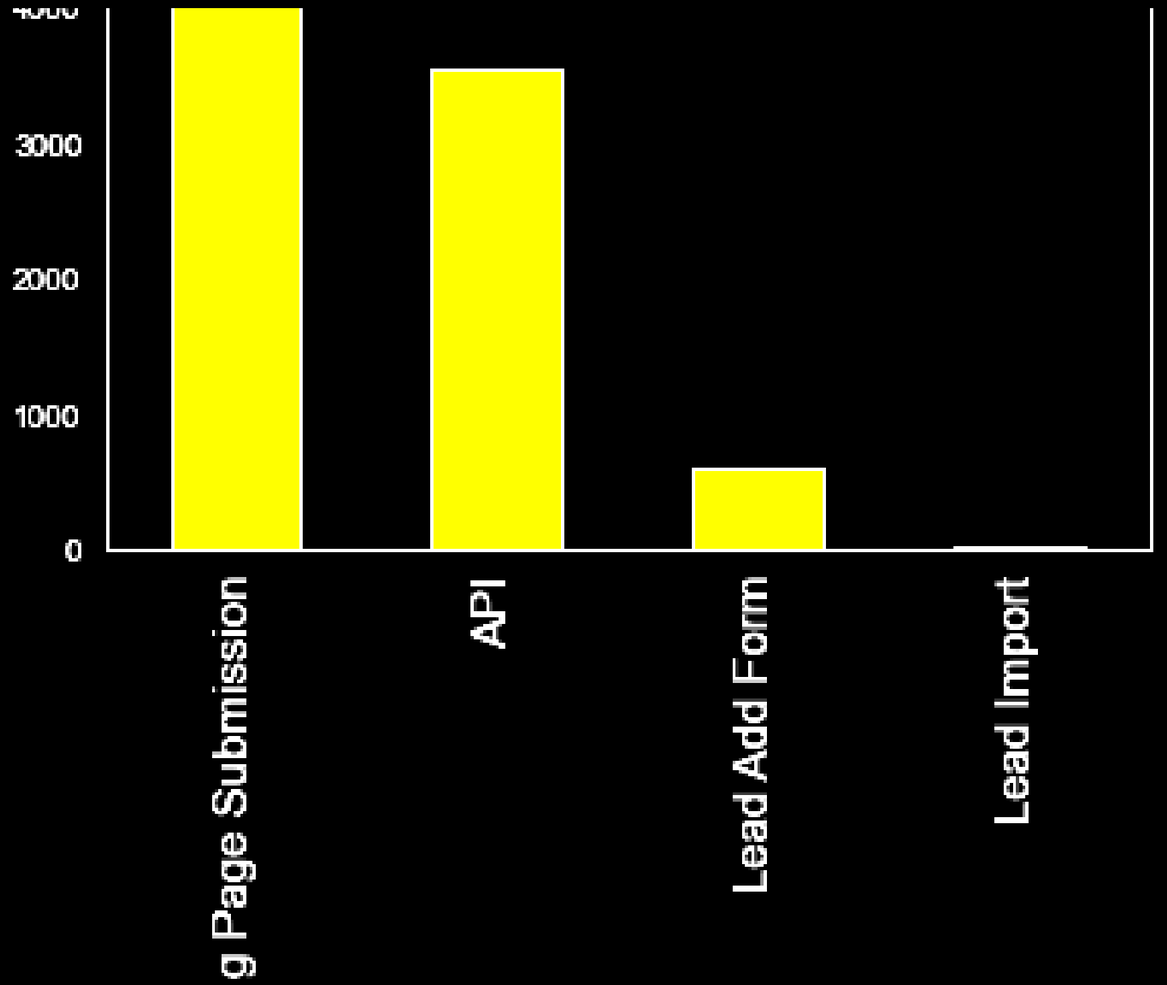
- Reading and Understanding the data : Importing and Cleaning the data.
- Visualizing Data : Exploratory data analysis for figuring out most helpful attributes
- Data Preparation : Creating dummy variables and Scaling features.
- Splitting the Data into Train and Test Set
- Model Building : Preparation of Logistic Regression Model
- Model Evaluation : Confusion Matrix, Accuracy, Sensitivity, Specificity and other metrics.
- Plotting the ROC Curve
- Finding the Optimal Cut off
- Making Predictions on Test Set
- Precision and Recall Analysis
- Making Predictions on Test Set Using the New Cut off

UNIVARIATE ANALYSIS

LEAD SOURCE

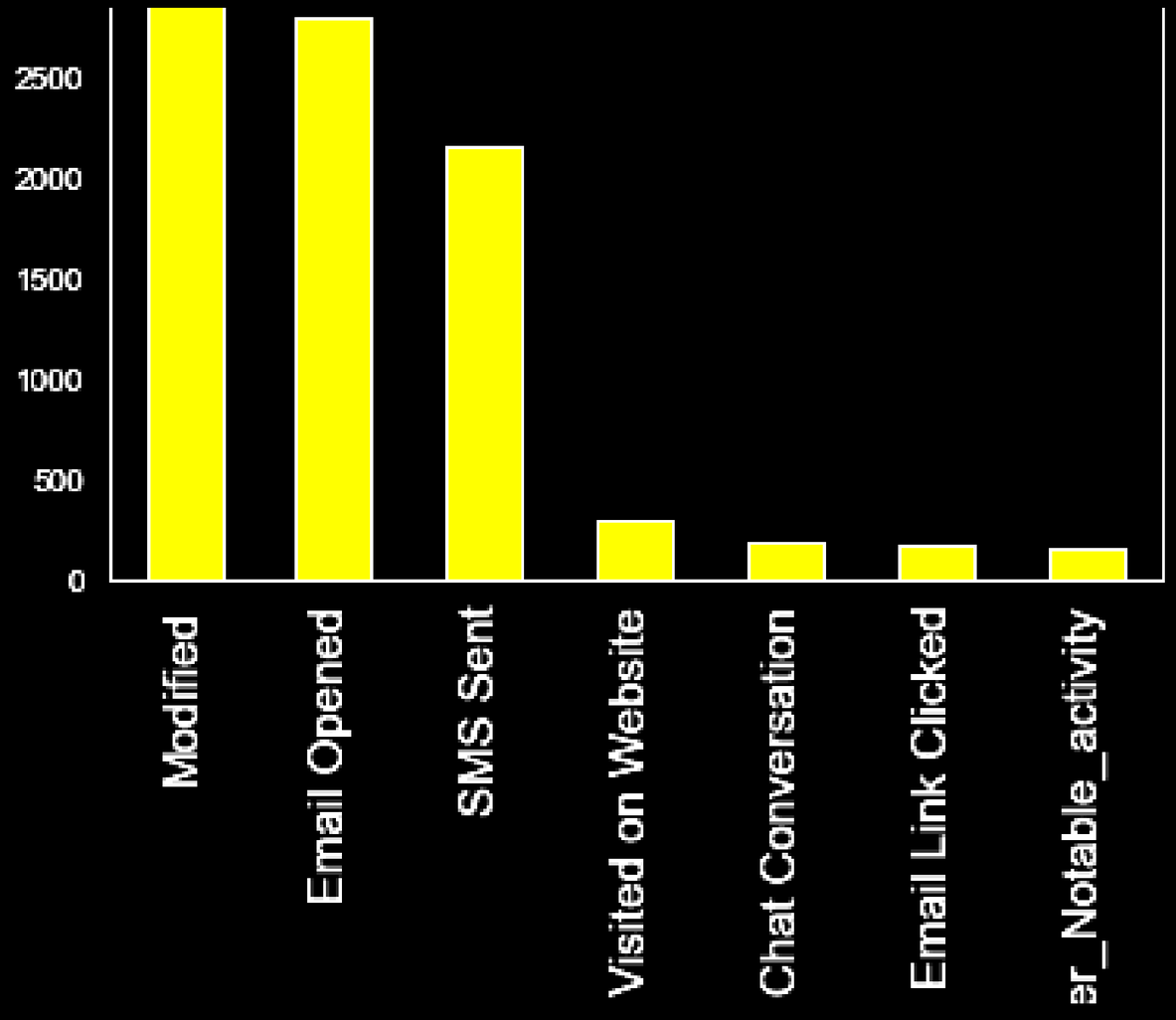


LEAD ORIGIN

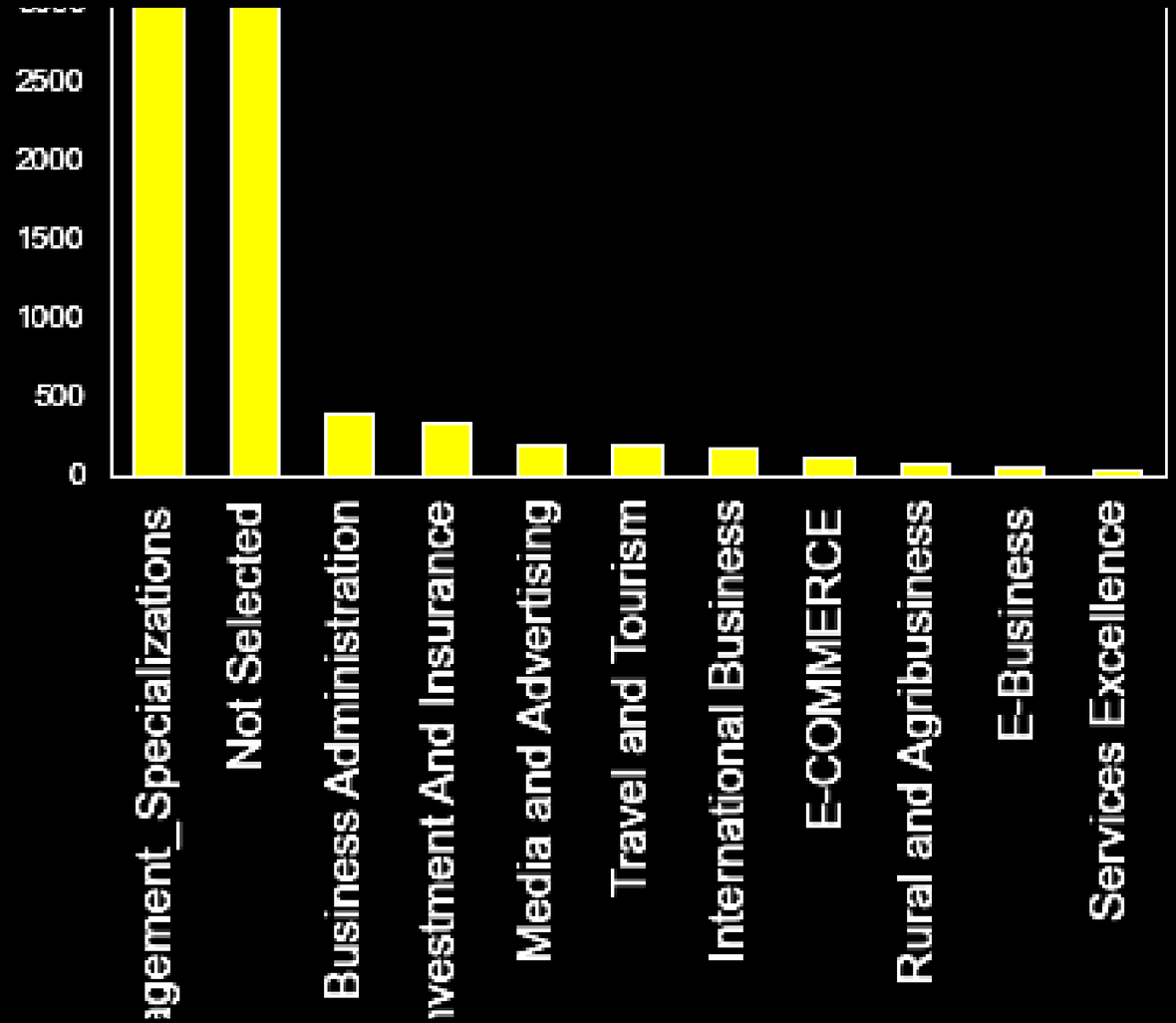


UNIVARIATE ANALYSIS

LAST NOTABLE ACTIVITY

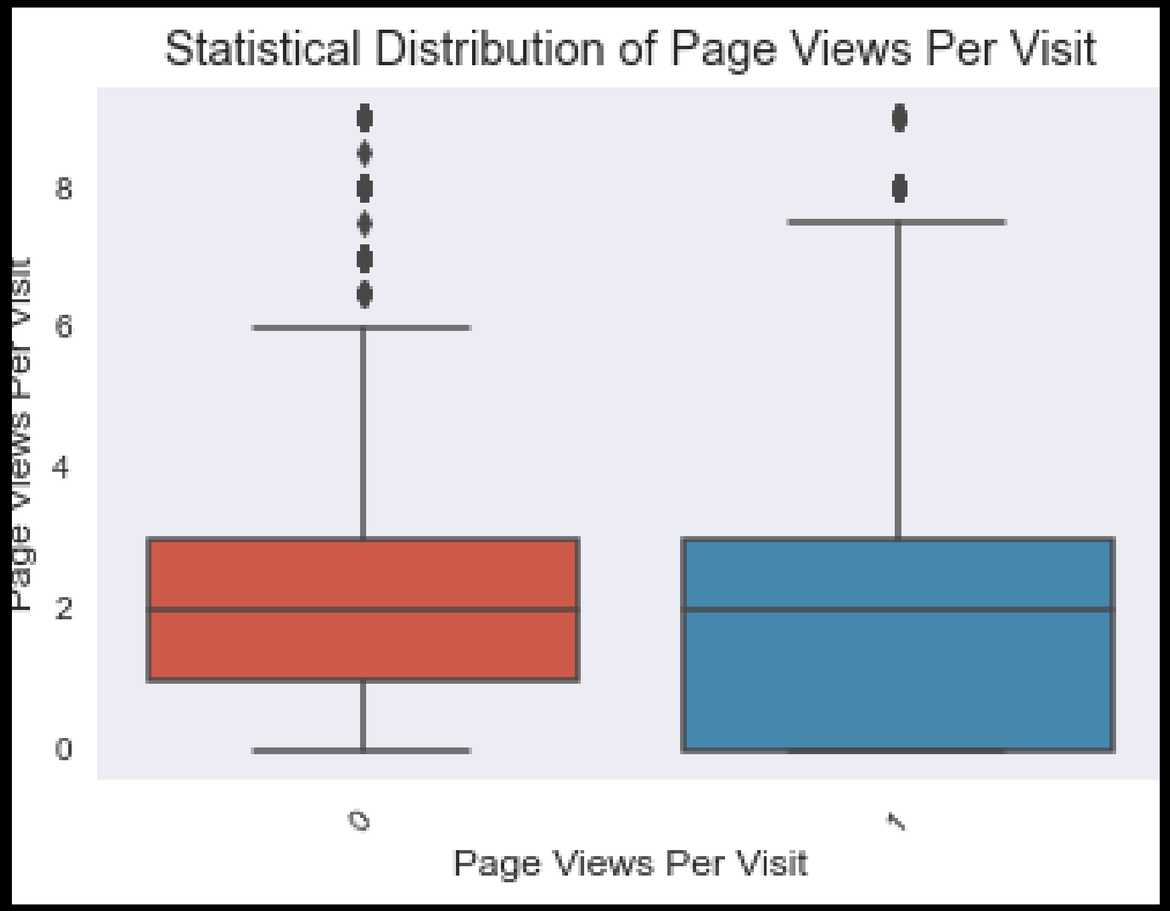


SPECIALIZATION

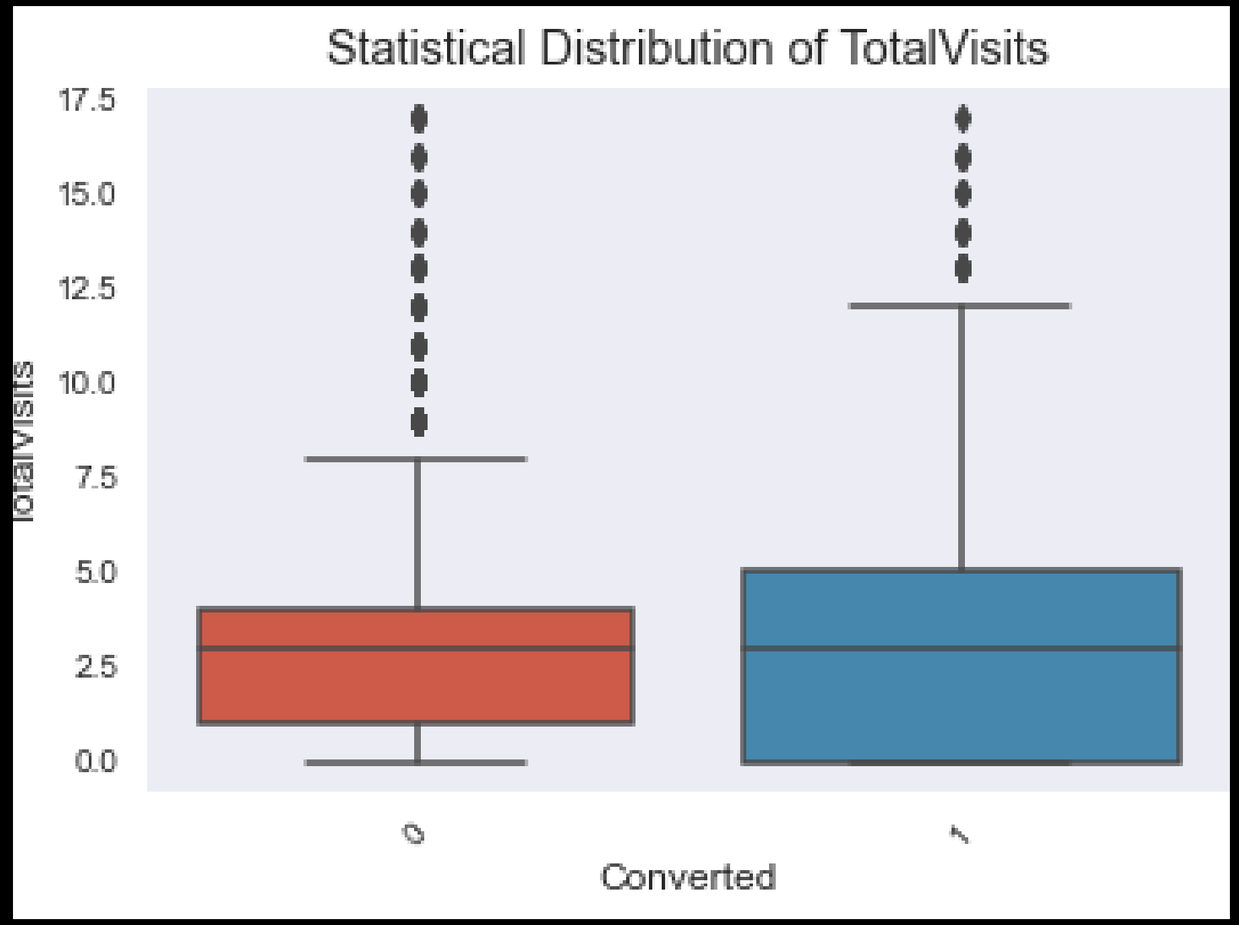


BIVARIATE ANALYSIS
HERE CONVERTED VARIABLE MENTION WITH 0 ‘NO’ AND 1 ‘YES’

PAGE VIEWS PER VISIT



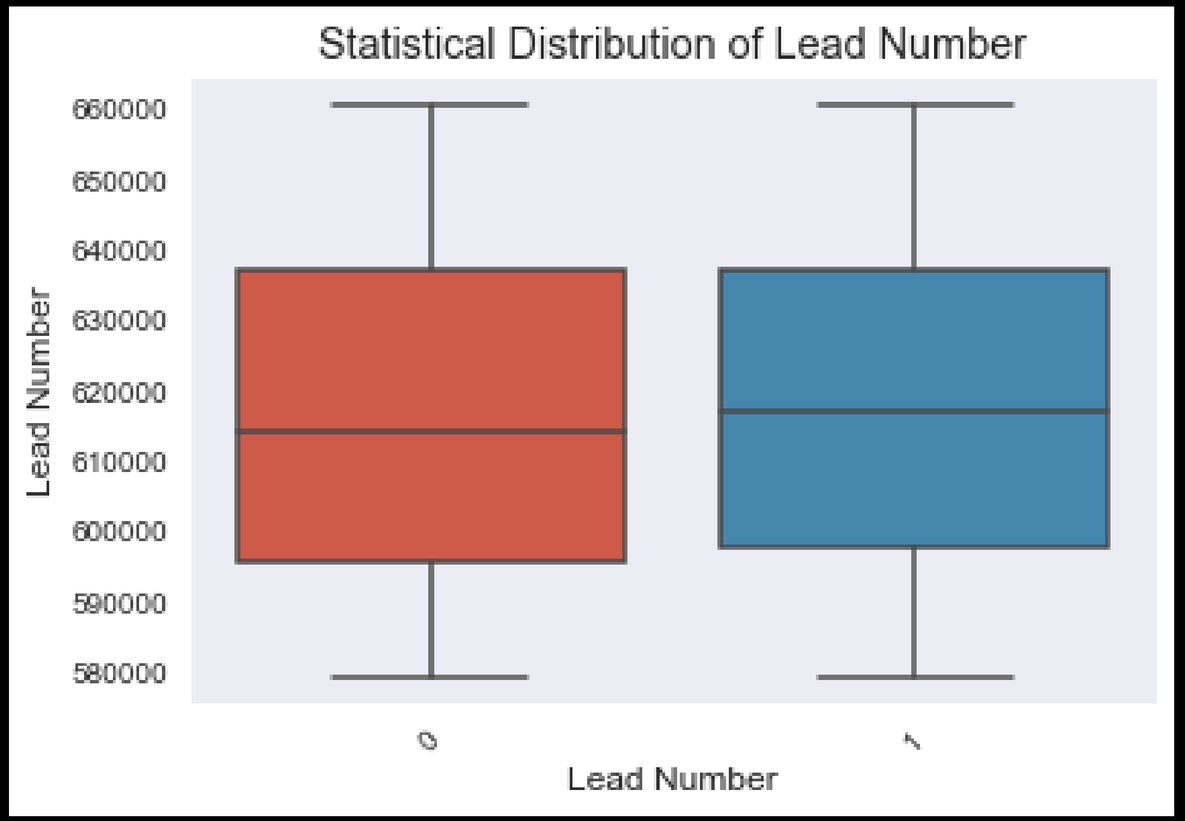
TOTAL VISITS



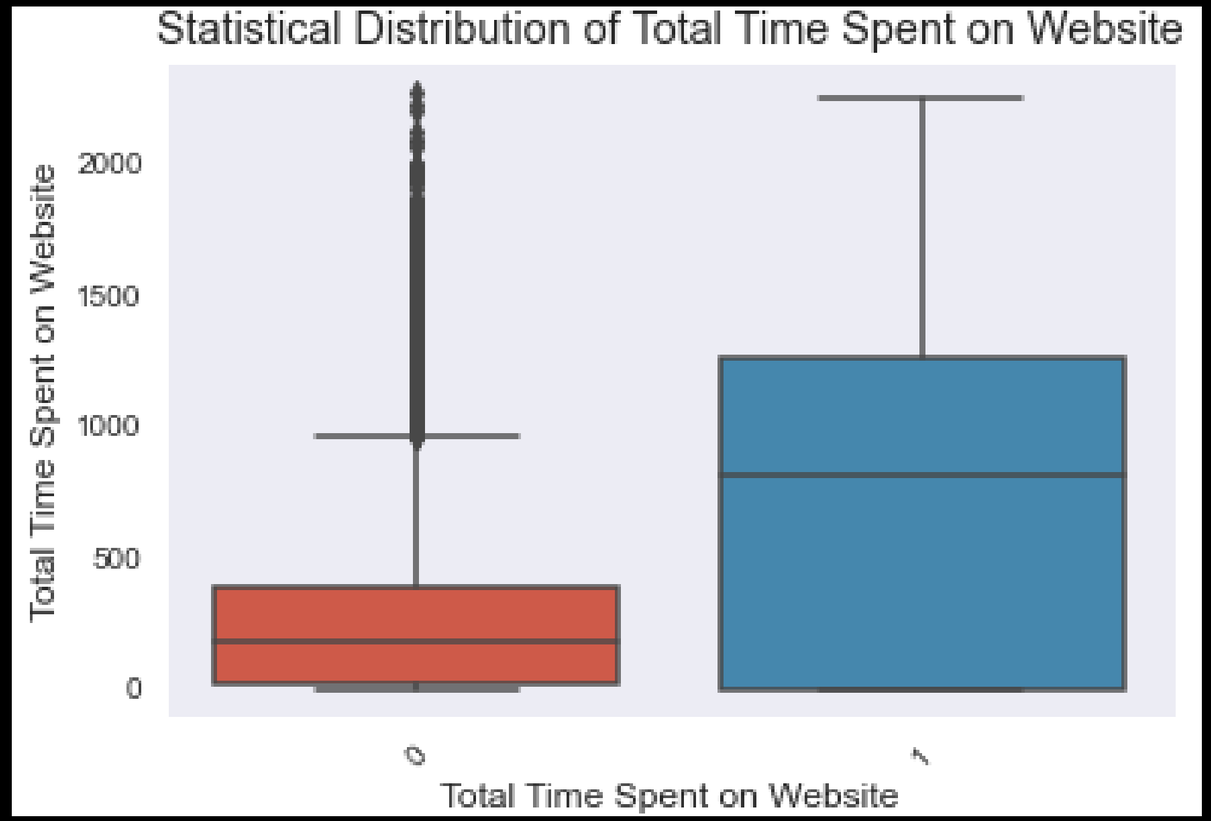
BIVARIATE ANALYSIS

HERE CONVERTED VARIABLE MENTION WITH O ‘NO’ AND 1 ‘YES’

LEAD NUMBER



TOTAL TIME SPENT ON WEBSITE

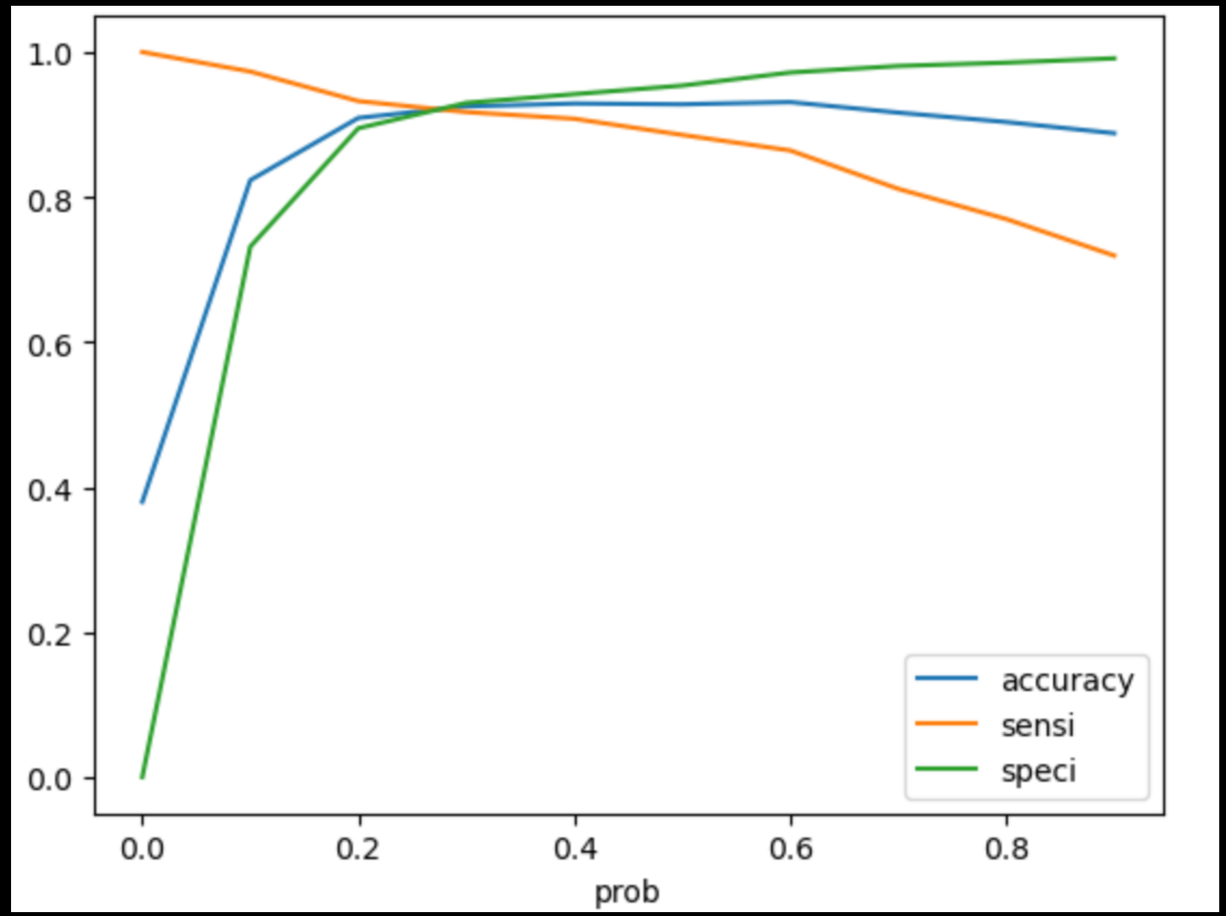


MODEL BUILDING

- Splitting the Data into Training and Testing Sets
- Rescaling the Features Here, we will use MinMax scaling on train set
- Build the first model
- Build the second model use RFE to eliminate less relevant variable
- Build the third model
- Eliminate variables based on the high p-values and high VIF
- Build the final model
- Metrics evaluation check confusion matrix, accuracy, sensitivity , specificity and other metrics
- Plotting the ROC curve
- Finding the optimal cut off point
- Making prediction on test set using the cut off
- Precision and RecallAnalysis
- Making prediction on test set with the new optimal cut off

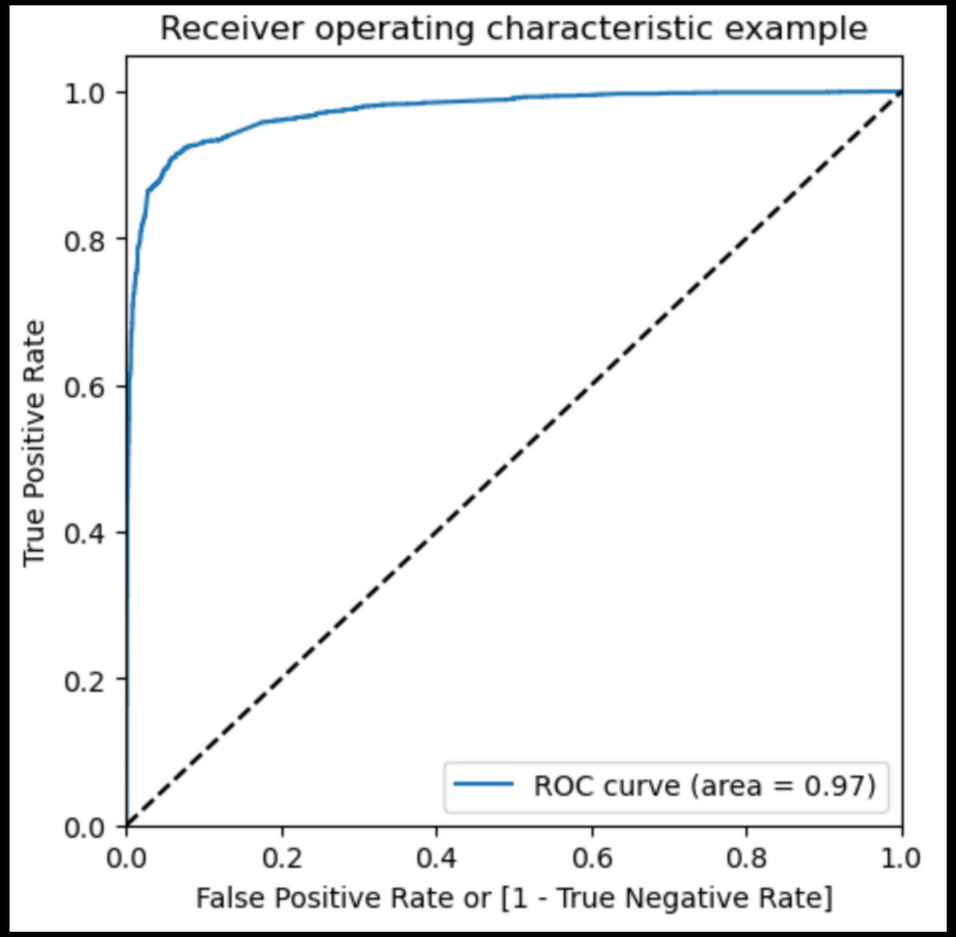
MODEL EVALUATION

- Accuracy : 92.79%
- Sensitivity : 88.57%
- Specificity : 95.37%



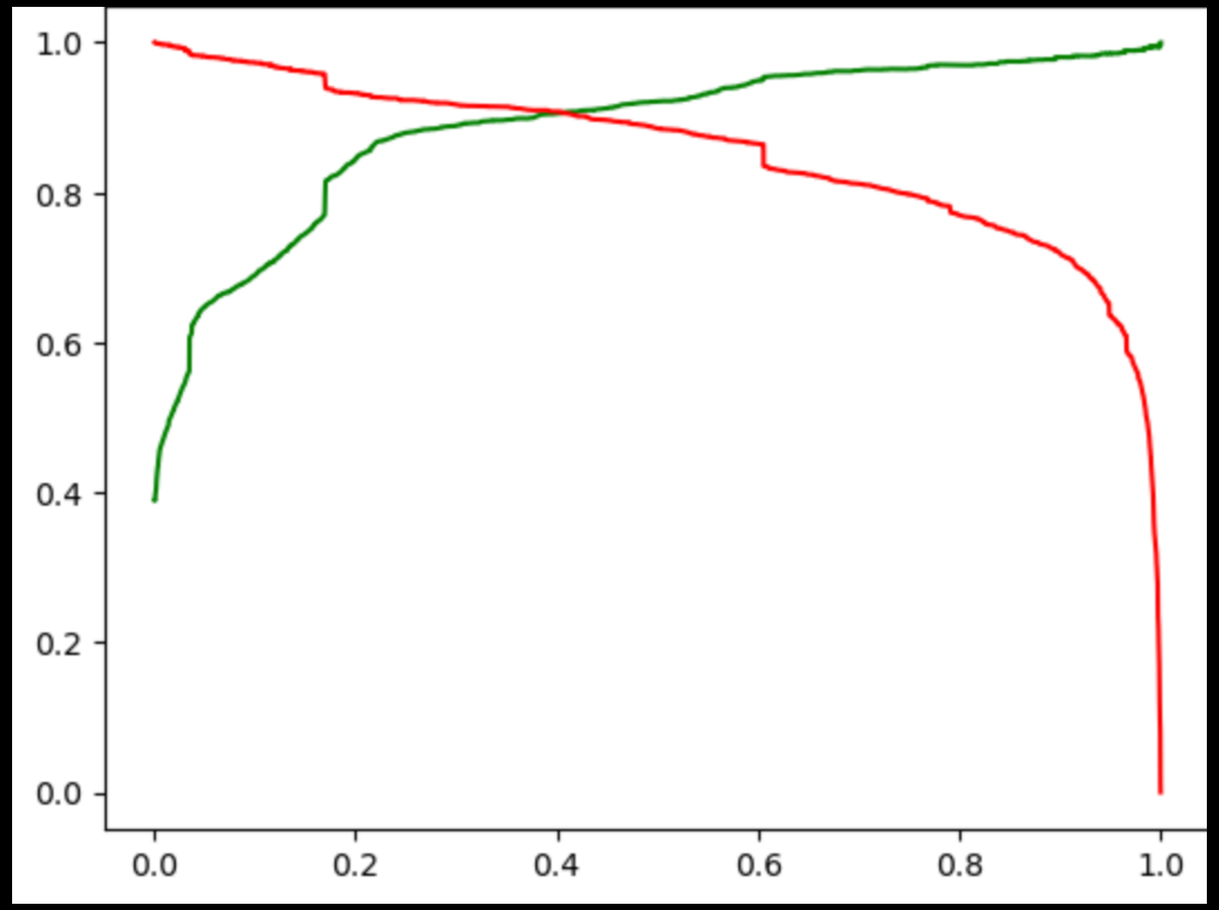
ROC CURVE

OBSERVATION: The area under the ROC curve is 0.97 which indicates a good model.



PRECISION - RECALL ANALYSIS

- With the present cutoff as 0.41 the Precision value is 87.03% and Recall value is 91.14%.



OBSERVATIONS ON TEST DATA:

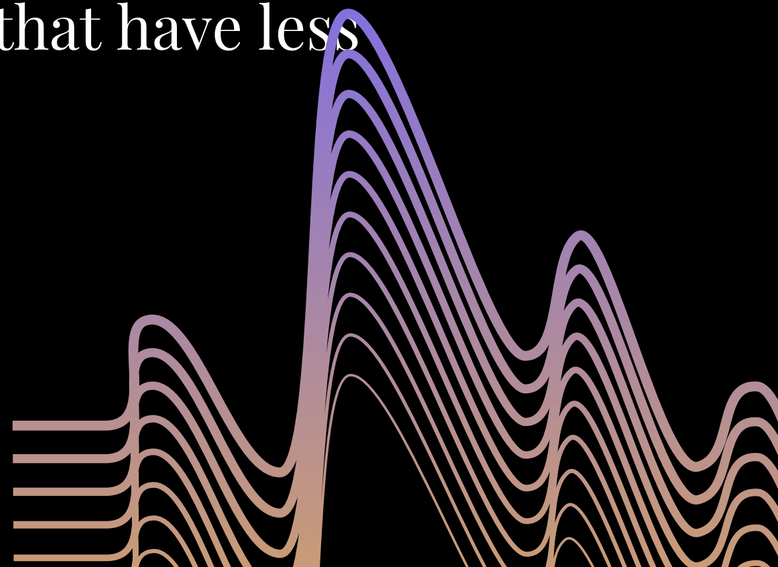
The analysis of the test data gives the following results:

- Accuracy : 92.18%
- Sensitivity : 89.96%
- Specificity : 93.53%

CONCLUSION

LOGISTIC REGRESSION MODEL

1. The model shows high close to 92 % accuracy
2. The model shows 88% sensitivity and 95% specificity
3. The threshold has been selected from accuracy, sensitivity, specificity measures and precision , recall curves.
4. The model finds correct promising leads and leads that have less chances of getting converted.
5. Overall this model can be said to be good.



RECOMMENDATIONS

According to the results of the model, phone calls could to be made to people who :

1. People who spend a lot of time on the website. The website needs to be made more insightful and interactive to get people to spend more time there, that will help in assisting lead conversion.
2. Making follow-up calls to people who are listed as Tags_Closed by Horizzon, Tags_Lost to EINS or Tags_Will revert after reading the email, as these people show a higher conversion rate according to the results of the model.
3. People whose last activity was through SMS are also good target customers as they have higher conversion potential.
4. Leads who have been obtained from Lead Origin ‘Lead Add Form’ are also potential customers since they have higher conversion rates than others.