LEADS CONVERTING PPT

PROBLEM STATEMENT

AN EDUCATION COMPANY NAMED X EDUCATION SELLS ONLINE COURSES TO INDUSTRY PROFESSIONALS. ON ANY GIVEN DAY, MANY PROFESSIONALS WHO ARE INTERESTED IN THE COURSES LAND ON THEIR WEBSITE AND BROWSE FOR COURSES.

THE COMPANY MARKETS ITS COURSES ON SEVERAL WEBSITES AND SEARCH ENGINES LIKE GOOGLE. ONCE THESE PEOPLE LAND ON THE WEBSITE, THEY MIGHT BROWSE THE COURSES OR FILL UP A FORM FOR THE COURSE OR WATCH SOME VIDEOS. WHEN THESE PEOPLE FILL UP A FORM PROVIDING THEIR EMAIL ADDRESS OR PHONE NUMBER, THEY ARE CLASSIFIED TO BE A LEAD. MOREOVER, THE COMPANY ALSO GETS LEADS THROUGH PAST REFERRALS. ONCE THESE LEADS ARE ACQUIRED, EMPLOYEES FROM THE SALES TEAM START MAKING CALLS, WRITING EMAILS, ETC. THROUGH THIS PROCESS, SOME OF THE LEADS GET CONVERTED WHILE MOST DO NOT. THE TYPICAL LEAD CONVERSION RATE AT X EDUCATION IS AROUND 30%.

NOW, ALTHOUGH X EDUCATION GETS A LOT OF LEADS, ITS LEAD CONVERSION RATE IS VERY POOR. TO MAKE THIS PROCESS MORE EFFICIENT, THE COMPANY WISHES TO IDENTIFY THE MOST POTENTIAL LEADS, ALSO KNOWN AS ‘HOT LEADS’. IF THEY SUCCESSFULLY IDENTIFY THIS SET OF LEADS, THE LEAD CONVERSION RATE SHOULD GO UP AS THE SALES TEAM WILL NOW BE FOCUSING MORE ON COMMUNICATING WITH THE POTENTIAL LEADS RATHER THAN MAKING CALLS TO EVERYONE

2. Higher scores signify hotter leads, indicating a higher likelihood of conversion, while lower scores suggest colder leads with lower conversion probabilities.
3. Utilize the lead scoring system to target potential leads effectively, focusing efforts on those with higher scores for increased conversion rates.
3.Utilize the lead scoring system to target potential leads effectively, focusing efforts on those with higher scores for increased conversion rates.

GOAL

- ❖ DEVELOP A PREDICTIVE MODEL AIMED AT ASSIGNING LEAD SCORES RANGING FROM 0 TO 100 TO POTENTIAL LEADS.
- ❖ HIGHER SCORES SIGNIFY HOTTER LEADS, INDICATING A HIGHER LIKELIHOOD OF CONVERSION, WHILE LOWER SCORES SUGGEST COLDER LEADS WITH LOWER CONVERSION PROBABILITIES.
- ❖ UTILIZE THE LEAD SCORING SYSTEM TO TARGET POTENTIAL LEADS EFFECTIVELY, FOCUSING EFFORTS ON THOSE WITH HIGHER SCORES FOR INCREASED CONVERSION RATES.

APPORACHES

- **DATA UNDERSTANDING:**

  Importing Data and Check Statistics.

- **DATA CLEANING:**

  Check missing values/checking outliers and fix those by checking their statistics.

- **EXPLORATORY ANALYSIS:**

  Uni-Variate, Bi-Variate, and Correlation or pair plots.

- **DATA PREPARATION:**

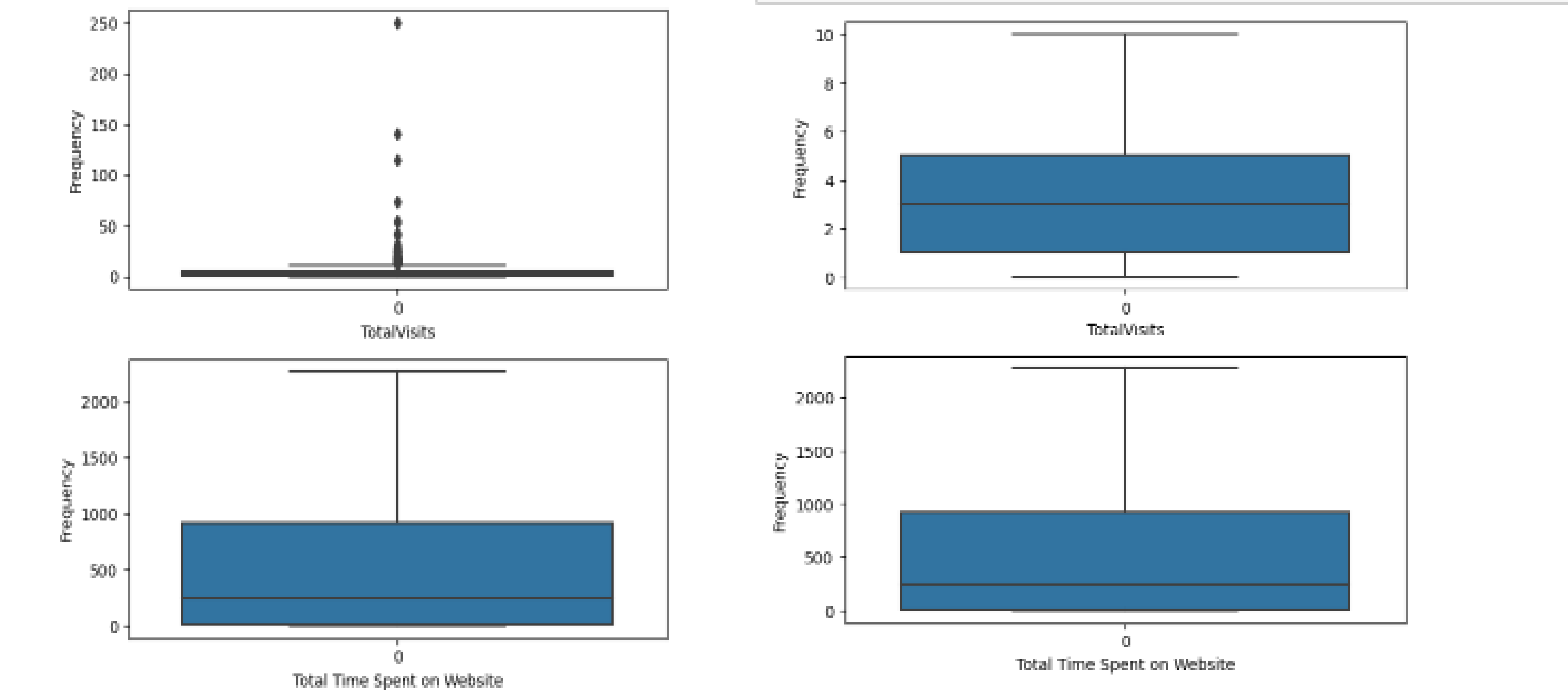
  Convert in binary column and dummy variables creation.

  Feature Scaling.

- **Build Model:**
  - Split the data into train and test sets.
  - Feature scaling.
  - Check correlation matrix.
- **Features Selection:**
  - Using RFE and manual methods.
- **Model Evaluation:**
  - Confusion Matrix.
  - Accuracy, specificity, Precision, and recall.
  - ROC Curve.
- **Prediction of Test Data:**
  - Apply the model to predict outcomes on the test data.

DATA UNDERSTANDING AND CLEANING

- START BY READING THE CSV FILE.
- CHECK THE SHAPE OF THE DATA TO UNDERSTAND ITS DIMENSIONS.
- EXAMINE INFORMATION AND DATA TYPES TO GAIN INSIGHTS INTO THE DATASET.
- IDENTIFY MISSING DATA AND COLUMNS WITH INACCURATE DATA TYPES.
- IMPUTE MISSING VALUES IN SELECTED COLUMNS WITH APPROPRIATE VALUES.
- REMOVE COLUMNS WITH A HIGH PERCENTAGE OF MISSING DATA TO STREAMLINE THE DATASET.
- DETECT OUTLIERS IN THE DATA AND ADDRESS THEM BY CAPPING THEIR VALUES.
- EXCLUDE COLUMNS LIKE "DO NOT EMAIL," "DO NOT CALL," AND OTHERS, WHICH CONTAIN MOSTLY ONE VARIABLE AND PROVIDE LITTLE MEANINGFUL INSIGHT.



• - As per boxplot, Totaltimr spent having outliers
• - After checking the percentile , find the interquartile range first and
• then cap on minimum and maximum values.

EXPLORATORY ANALYSIS

Slide: Managing Outliers in Website Metrics

- Title: Managing Outliers in Website Metrics
- Introduction:
  - Outliers in metrics like Total Visits and Page Views per Visit can distort data analysis.
- Method:
  - Use boxplots to visualize data distribution.
  - Calculate percentiles to determine outlier thresholds.
  - Apply Interquartile Range (IQR) to cap outliers.
- Action Steps:
  - Identify outliers in Total Visits and Page Views per Visit using boxplots.
  - Calculate the IQR for both metrics.
  - Set minimum and maximum thresholds based on IQR.
  - Replace outlier values beyond these thresholds.
- Benefits:
  - Ensures more accurate data analysis and interpretation.
  - Improves the reliability of insights derived from website metrics.
- Conclusion:
  - Managing outliers is essential for robust data analysis and decision-making in website performance evaluation.

INFERENCES FROM UNI AND BI-VARIATE ANALYSIS

SLIDE: KEY INSIGHTS FROM CONVERTED LEADS ANALYSIS

TITLE: KEY INSIGHTS FROM CONVERTED LEADS ANALYSIS

LEAD ORIGIN AND SOURCE:

Most converted leads originate from Landing Page submissions.

Google is the primary source of converted leads.

LAST ACTIVITY:

SMS Sent is the most common last activity for converted leads.

SPECIALIZATION:

Converted leads tend not to choose a specialization, indicating either unavailability or lack of interest in the options provided.

ONLINE SEARCH:

Majority of converted leads search online for X Education.

EMPLOYMENT STATUS:

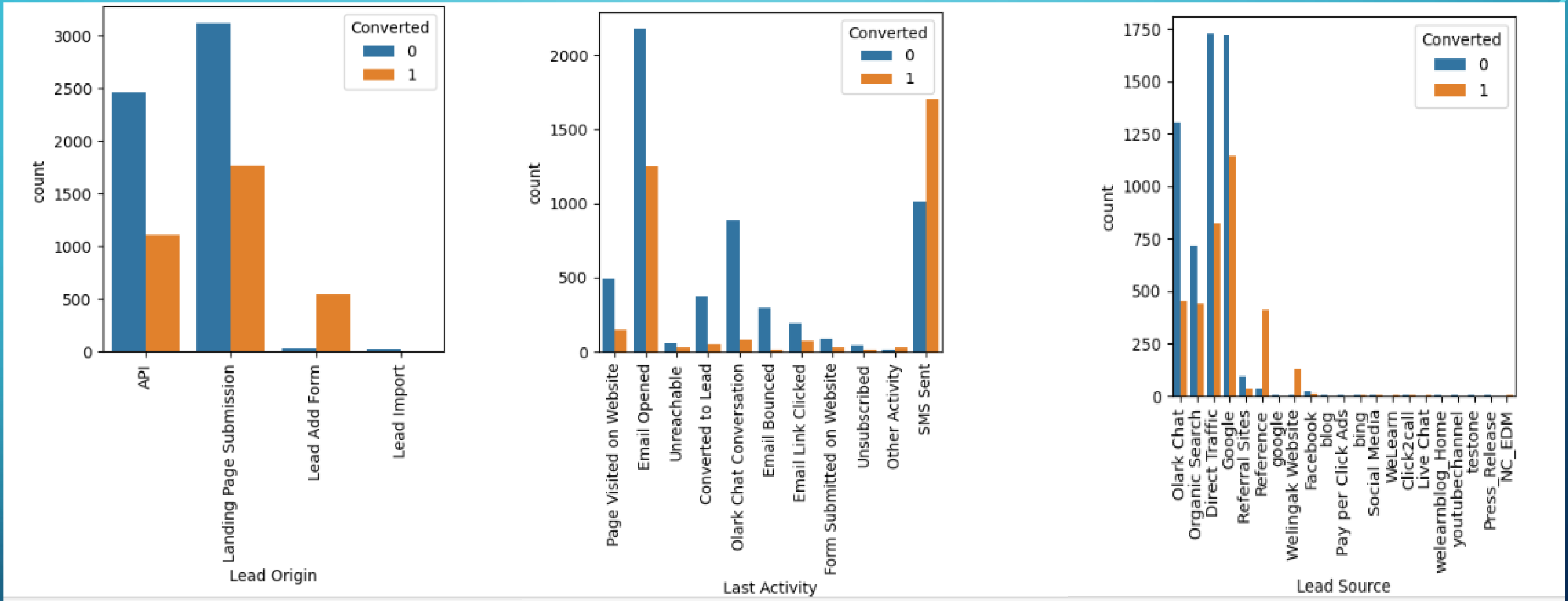
Unemployed individuals are the most common among converted leads.

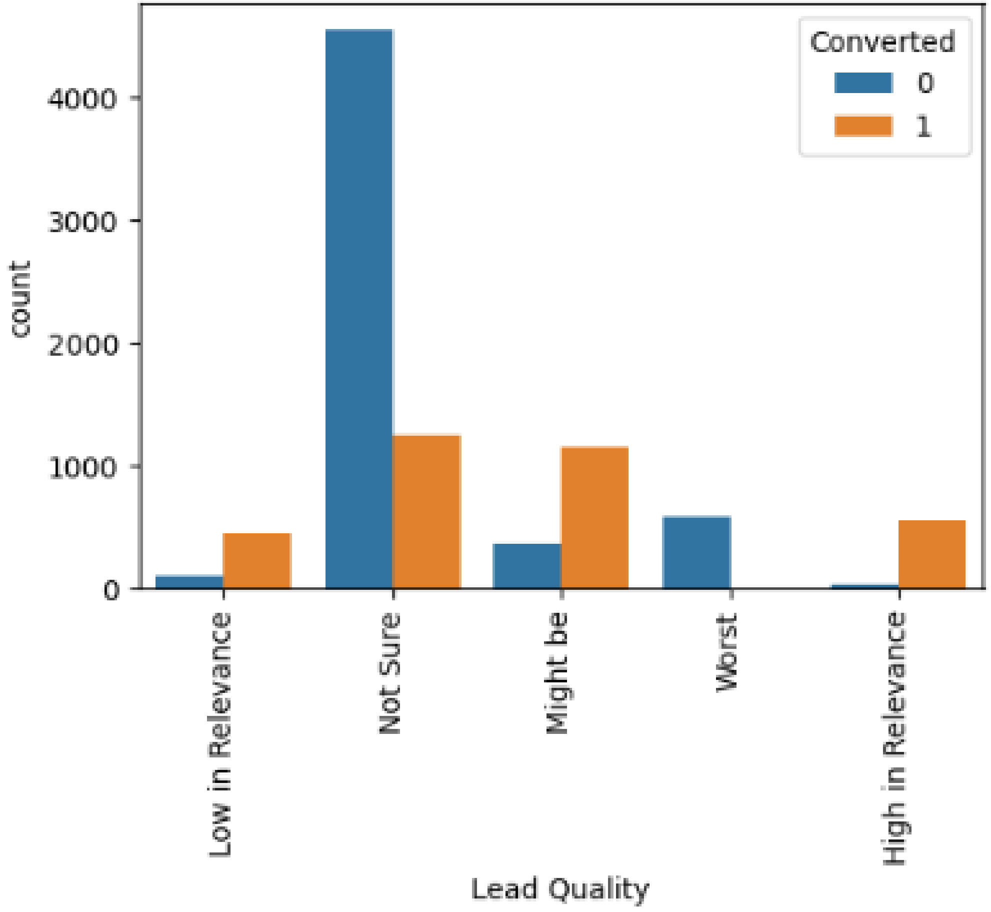
TAGS MANAGEMENT:

Tags revert after reading emails with the highest conversion rate.

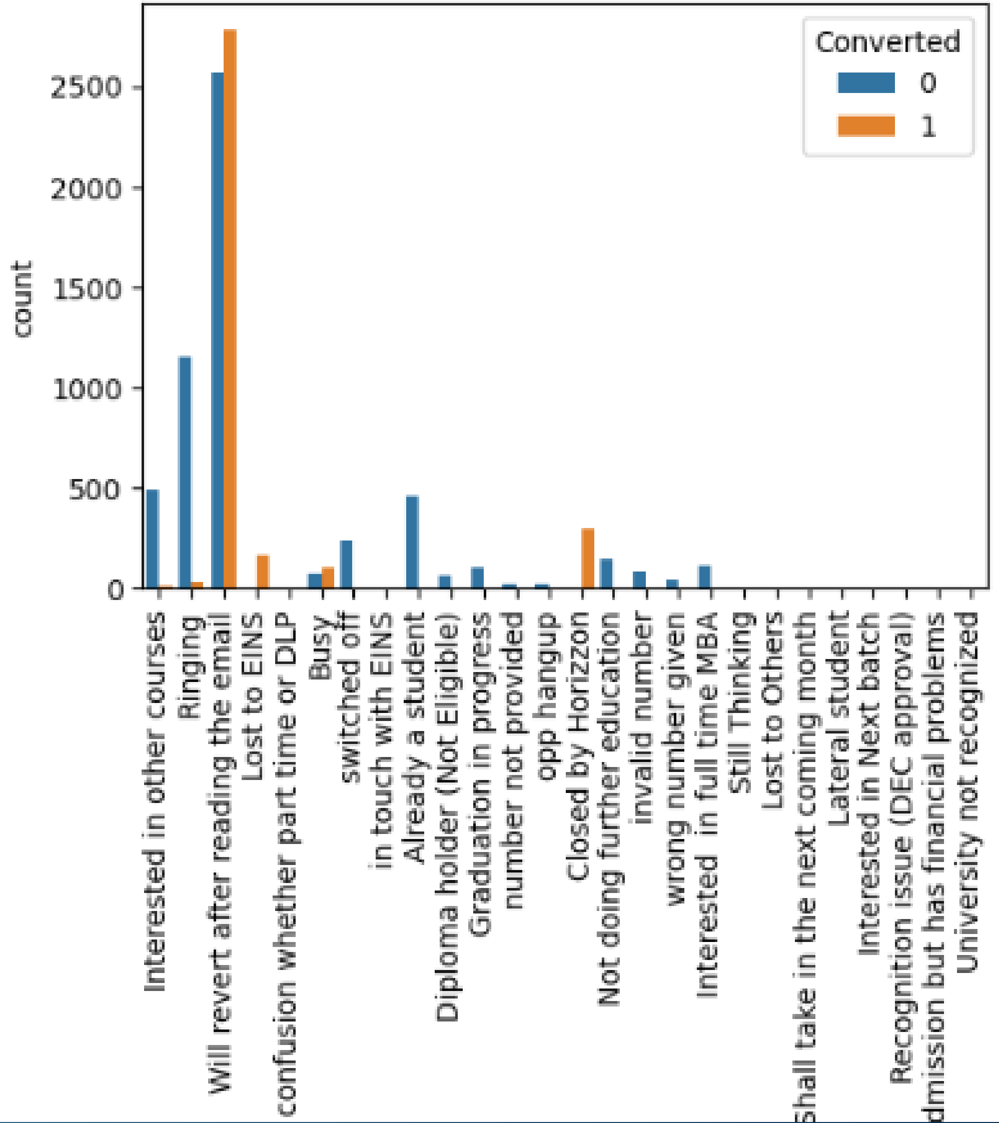
1. Lead Profile:
   - Potential leads are identified based on lead profiles.
2. Location:
   - Mumbai City is the top location for converted leads.
3. Last Activity:
   - SMS Sent is the predominant last activity for converted leads.
4. Statistical Analysis:
   - The 50th percentile values for Lead Number, Total Visits, and Page Views per Visit are similar for converted and non-converted leads, limiting inference potential.
5. Total Time Spent:
   - Total time spent on the website is higher for converted leads.

- Conclusion: Understanding these insights can help optimize marketing strategies and improve lead conversion rates.

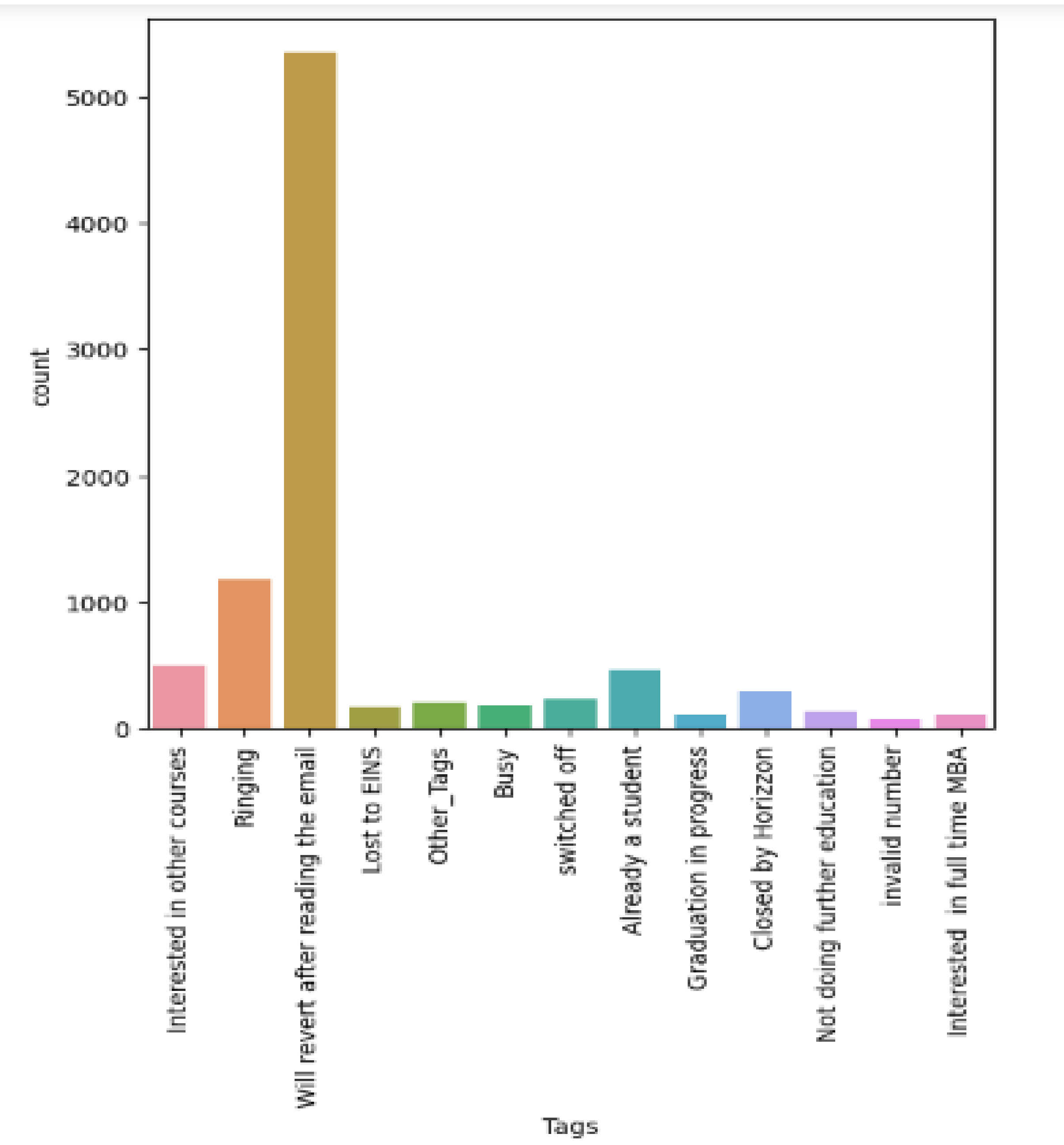

count
3000
2500
2000
1500
1000
500
0
Converted
0
1
API
Landing Page Submission
Lead Add Form
Lead Import
Lead Origin
count
2000
1500
1000
500
0
Converted
0
1
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Other Activity
SMS Sent
Last Activity
count
1750
1500
1250
1000
750
500
250
0
Converted
0
1
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Reference
google
Welingak Website
Facebook
blog
Pay per Click Ads
bing
Social Media
WeLearn
Click2call
Live Chat
welearnblog_Home
youtubechannel
testone
Press_Release
NC_EDM
Lead Source



count
4000
3000
2000
1000
0
Converted
0
1
Low in Relevance
Not Sure
Might be
Worst
High in Relevance
Lead Quality



count
2500
2000
1500
1000
500
0
Converted
0
1
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
confusion whether part time or DLP
Busy
switched off
in touch with EINS
Already a student
Diploma holder (Not Eligible)
Graduation in progress
number not provided
opp hangup
Closed by Horizzon
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Still Thinking
Lost to Others
Shall take in the next coming month
Lateral student
Interested in Next batch
Recognition issue (DEC approval)
dmission but has financial problems
University not recognized


count
5000
4000
3000
2000
1000
0
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
Other_Tags
Busy
switched off
Already a student
Graduation in progress
Closed by Horizzon
Not doing further education
invalid number
Interested in full time MBA
Tags

DATA PREPARATION

- CONVERT TARGET: CHANGE "CONVERTED" COLUMN TO BINARY (0/1) FOR CONVERSION STATUS.
- DUMMY VARIABLES: CREATE BINARY INDICATORS FOR CATEGORICAL COLUMNS LIKE "DO NOT EMAIL," "DO NOT CALL," ETC.
- SCALING: STANDARDIZE FEATURE RANGES FOR UNIFORMITY USING METHODS LIKE MIN-MAX SCALING OR STANDARDIZATION.
- BENEFITS: FACILITATES ALGORITHM CONVERGENCE, FAIR FEATURE COMPARISON, AND IMPROVES MODEL PERFORMANCE.
- IMPLEMENTATION: UTILIZE LIBRARIES LIKE SCIKIT-LEARN IN PYTHON FOR EFFICIENT PREPROCESSING.
- QUALITY CHECK: VALIDATE TRANSFORMATIONS FOR ACCURACY AND INTEGRITY, ENSURING PROPER ENCODING AND ANOMALY DETECTION.
- DOCUMENTATION: THOROUGHLY DOCUMENT PREPROCESSING STEPS FOR REPRODUCIBILITY AND TRANSPARENCY.
- CONCLUSION: PROPER DATA PREPARATION ENSURES CLEAN, STANDARDIZED DATA, CRUCIAL FOR ACCURATE LEAD CONVERSION ANALYSIS.

MODEL BUILDING

RFE & LOGISTIC REGRESSION:

Use RFE for iterative feature selection in logistic regression modeling.

VARIABLE SIGNIFICANCE:

Stop when all variables have $p < 0.05$, indicating significance for lead conversion.

VIF MONITORING:

Check VIF to detect multicollinearity issues during iteration.

SELECTED FEATURES FOR CONVERSIONS:

Include features such as origin, source, activities, education, occupation, tags, and profile.

OPTIMIZATION:

Halt iteration when all variables are significant and VIF is low.

CONCLUSION:

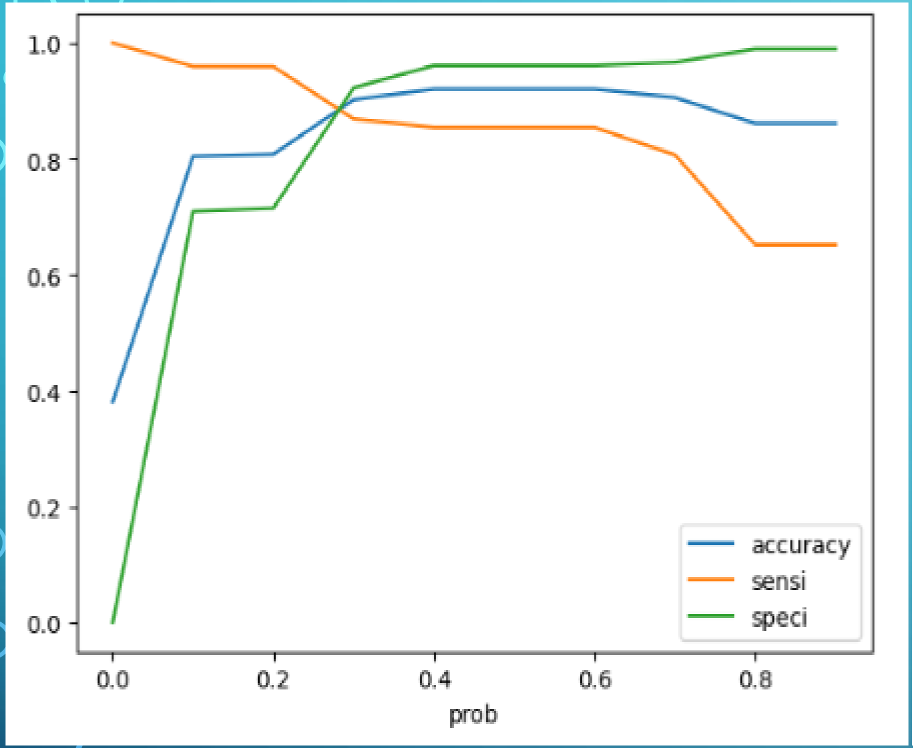
This approach ensures a concise logistic regression model tailored for predicting lead conversions.

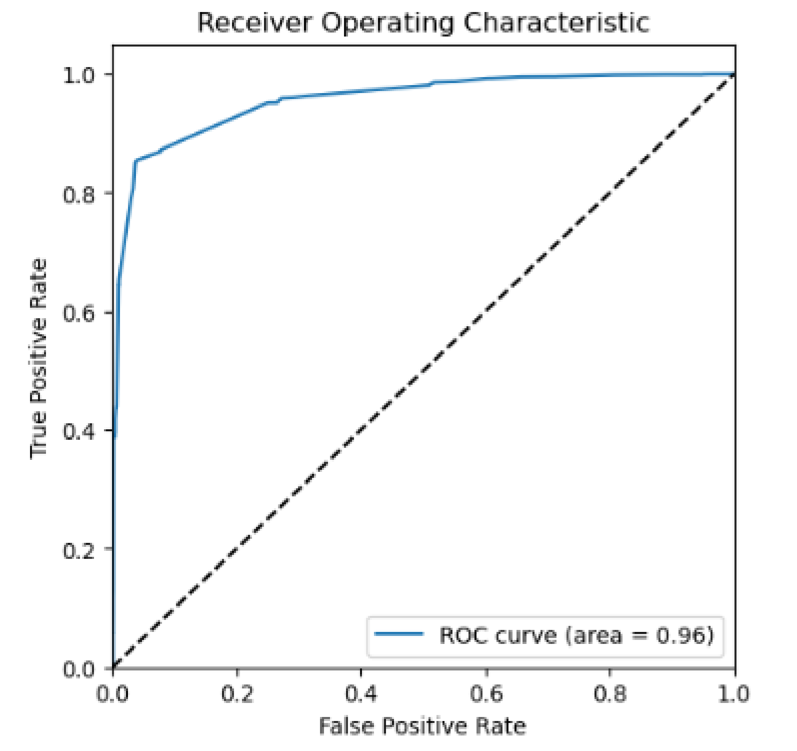
RUN MODEL

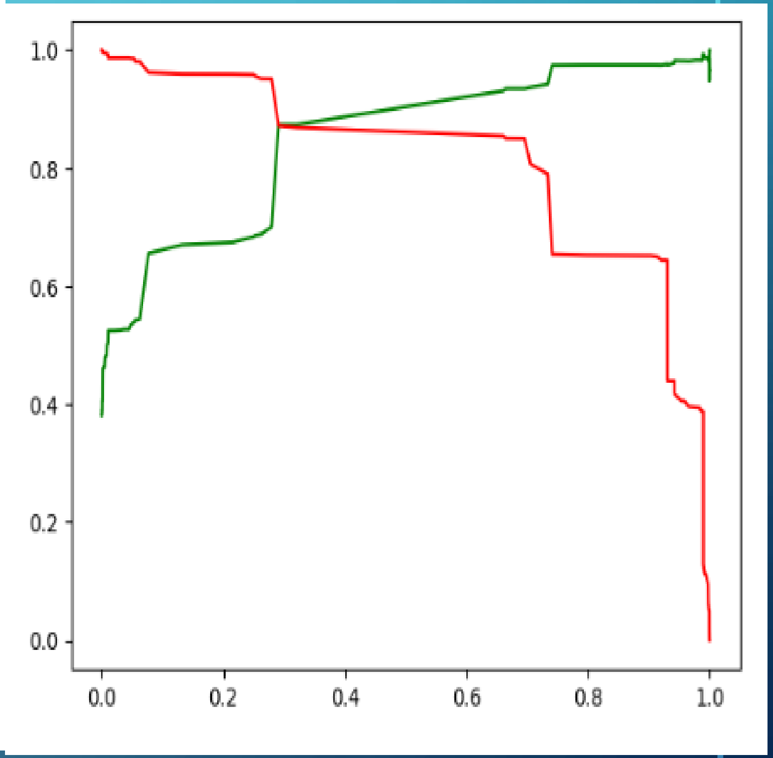
- GET THE PREDICTED VALUE ON TRAIN SET
- CREATED A DATA FRAME WITH ACTUAL CONVERTED AND CONVERTED PROBABILITY
- PROBABILITY WITH 50 % ABOVE ARE CONSIDER AS CONVERTED PREDICTED LEAD
- CALCULATE CONFUSION MATRIX

MODEL EVALUATION

- ON TRAIN DATA
- ACCURACY- 80%
- SENSITIVITY- 84%
- SPECIFICITY- 91%
- PRECISION- 85%
- RECALL-84%
- OPTIMAL CUT-OFF- 0.3
- - ON TEST DATA
- ACCURACY-85%
- SENSITIVITY- 84%
- SPECIFICITY- 85%


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
prob
accuracy
sensi
speci


Receiver Operating Characteristic
True Positive Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
ROC curve (area = 0.96)


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

THANK YOU