LEAD SCORE CASE STUDY

PREPARED BY :-
TANUSREE HALDER
JAI BATRA
THOMAS KURUVILA

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- 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.

- **Business Objective**:
- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

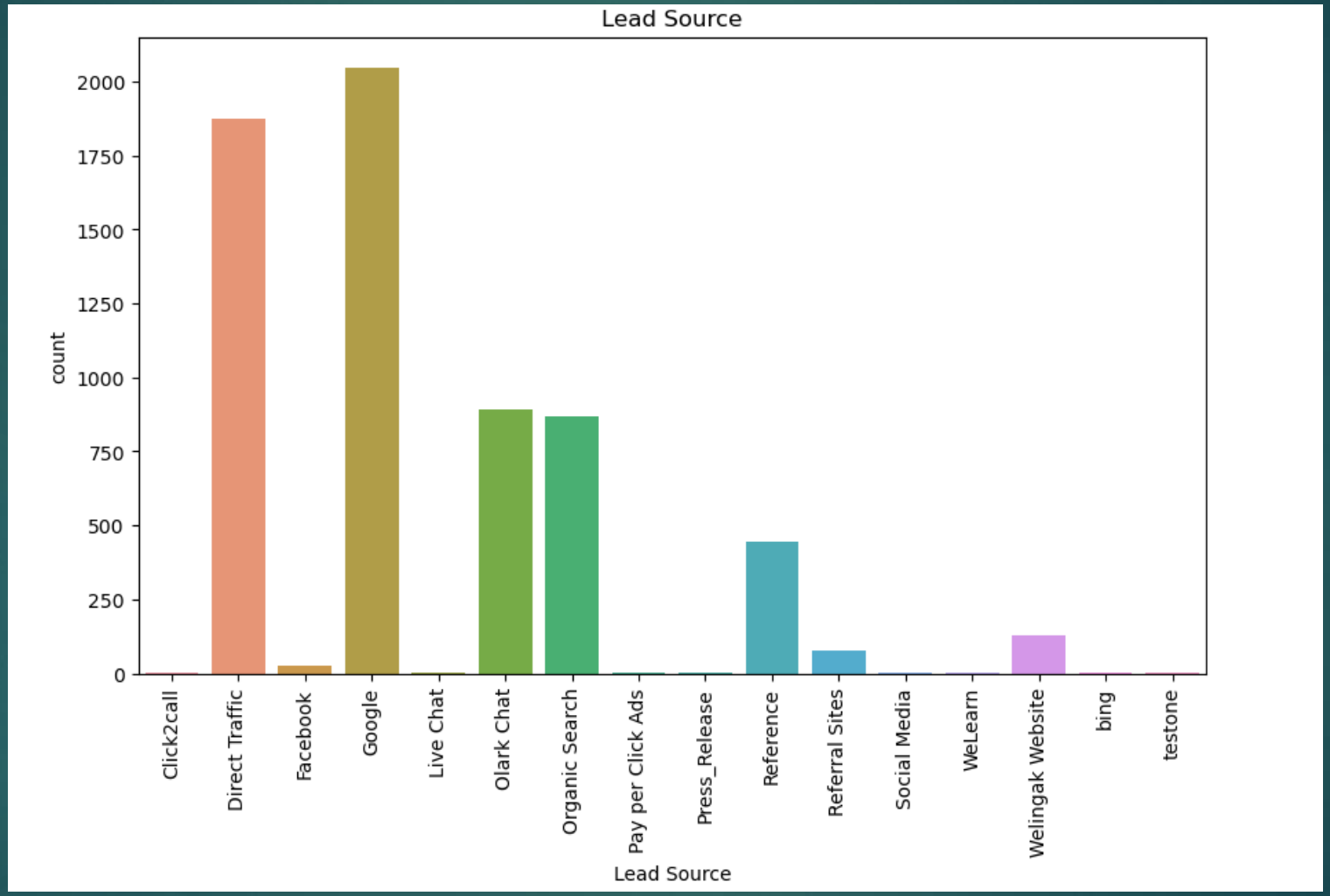
Solution Methodology

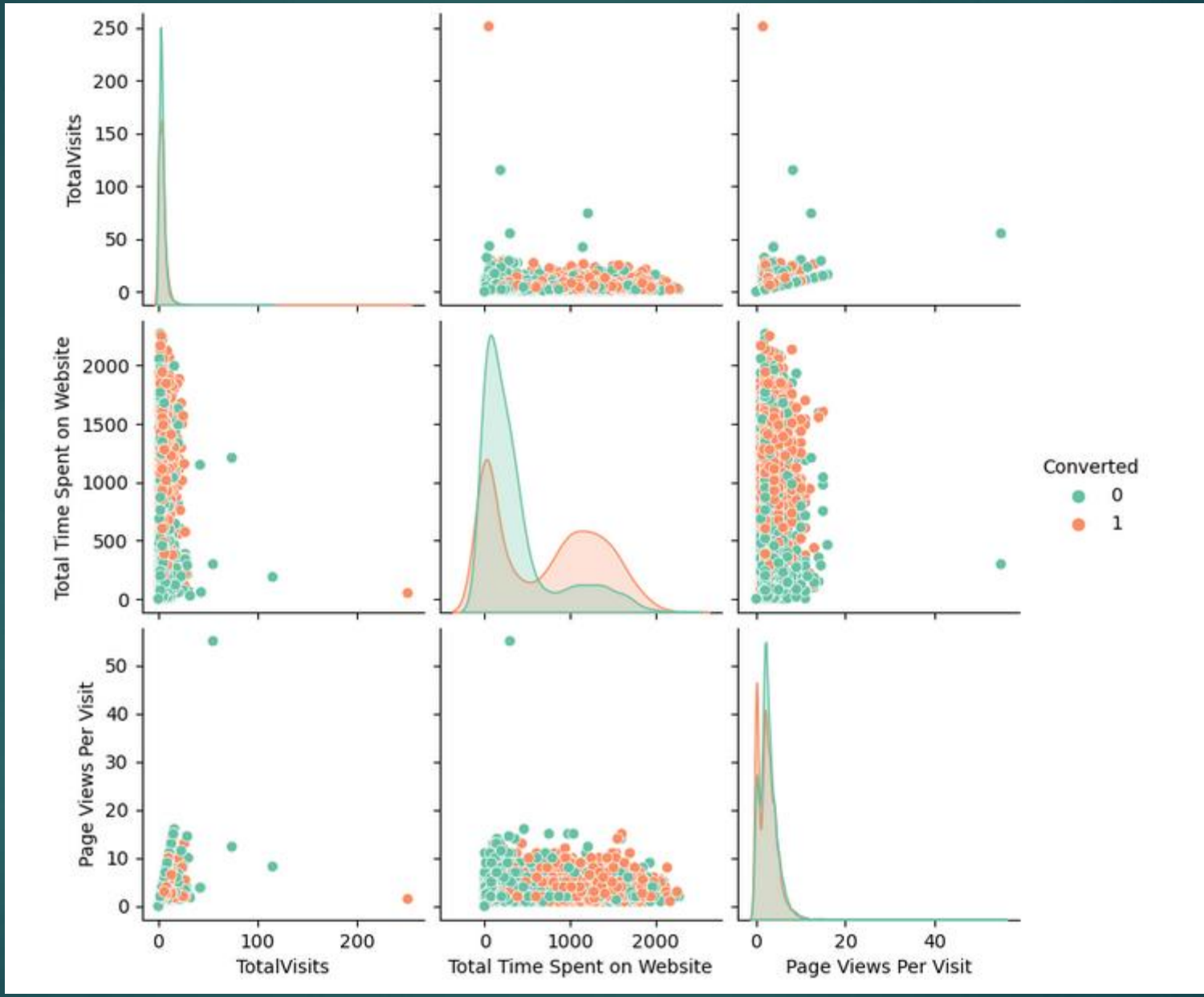
- Data Cleaning and Data Manipulation
- Check and handle duplicate data
- Check and handle NA values and missing values
- Drop columns with a large amount of missing values that are not useful for analysis
- Imputation of values if necessary
- Check and handle outliers in data
- Exploratory Data Analysis (EDA)
- Univariate data analysis: value count, distribution of variables, etc.
- Bivariate data analysis: correlation coefficients and patterns between variables, etc.
- Feature Scaling & creation of Dummy Variables and data encoding
- Classification technique: Logistic Regression for model building and prediction
- Model validation process
- Presentation of the model
- Conclusions and recommendations

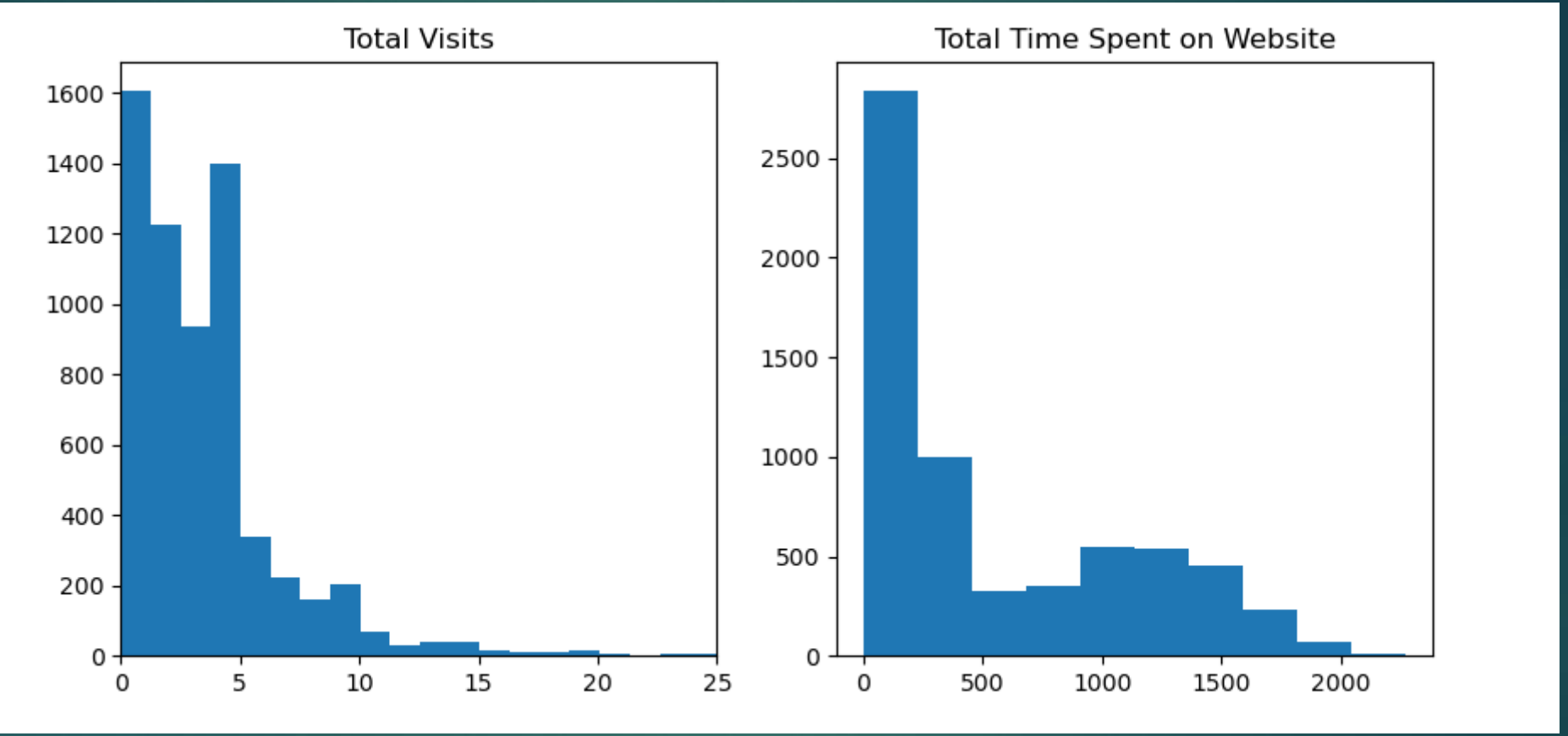
Data Manipulation

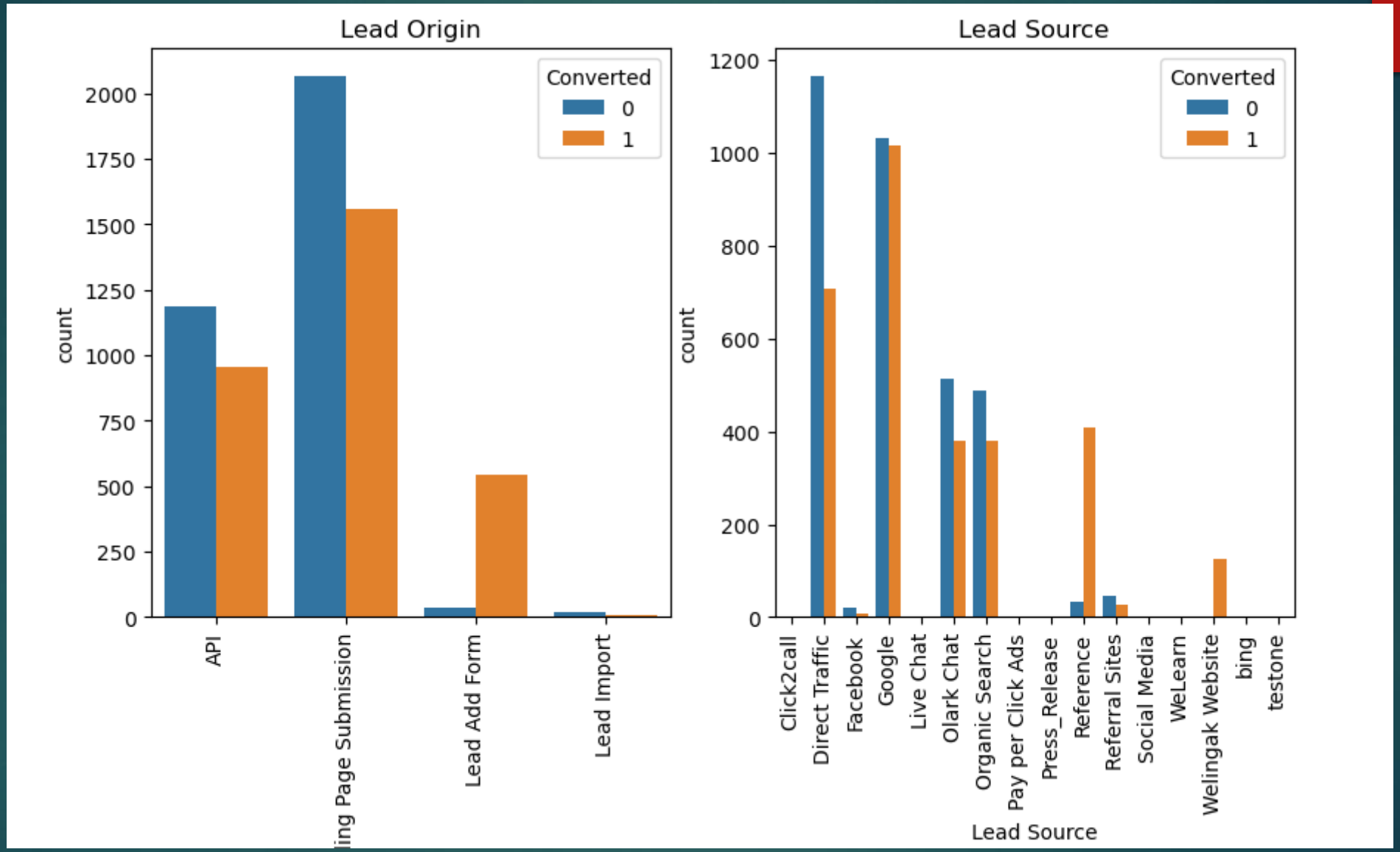
- Total number of rows: 37
- Total number of columns: 9240
- Dropped single value features:
  - "Magazine"
  - "Receive More Updates About Our Courses"
  - "Update me on Supply Chain Content"
  - "Get updates on DM Content"
  - "I agree to pay the amount through cheque"
- Removed "Prospect ID" and "Lead Number" as they are unnecessary for analysis
- Checked value counts for some object type variables
- Identified certain features with insufficient variance

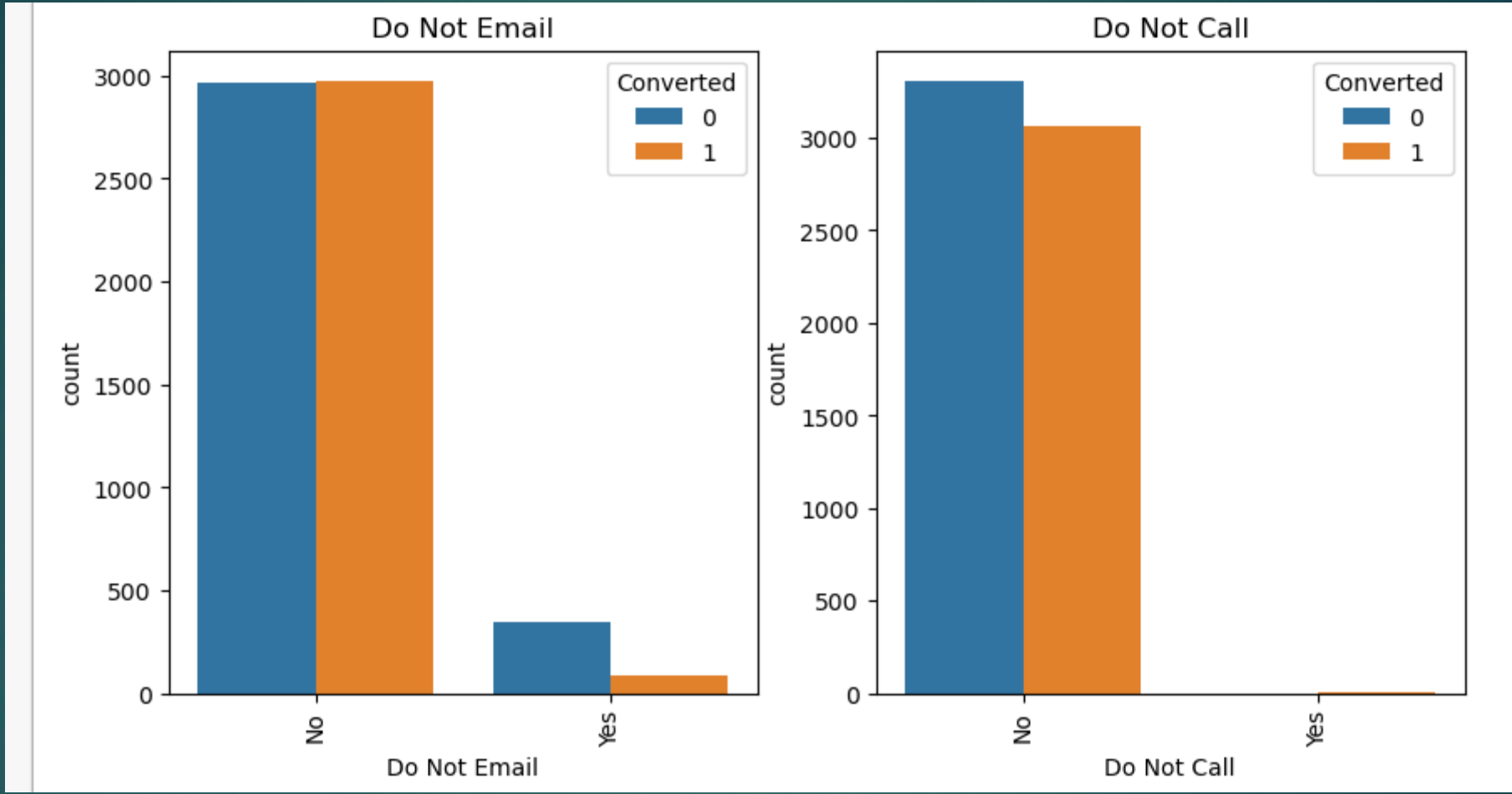
EDA


Lead Source
count
2000
1750
1500
1250
1000
750
500
250
0
Click2call
Direct Traffic
Facebook
Google
Live Chat
Olark Chat
Organic Search
Pay per Click Ads
Press_Release
Reference
Referral Sites
Social Media
WeLearn
Welingak Website
bing
testone
Lead Source


TotalVisits
Total Time Spent on Website
Page Views Per Visit
Converted
0
1
0
100
200
0
1000
2000
0
20
40


Total Visits
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
Total Time Spent on Website
2500
2000
1500
1000
500
0
0
500
1000
1500
2000


Lead Origin
Converted
0
1
count
2000
1750
1500
1250
1000
750
500
250
0
API
ing Page Submission
Lead Add Form
Lead Import
Lead Source
Converted
0
1
count
1200
1000
800
600
400
200
0
Click2call
Direct Traffic
Facebook
Google
Live Chat
Olark Chat
Organic Search
Pay per Click Ads
Press_Release
Reference
Referral Sites
Social Media
WeLearn
Welingak Website
bing
testone
Lead Source


Do Not Email
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
No
Yes
Do Not Email
Do Not Call
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
No
Yes
Do Not Call

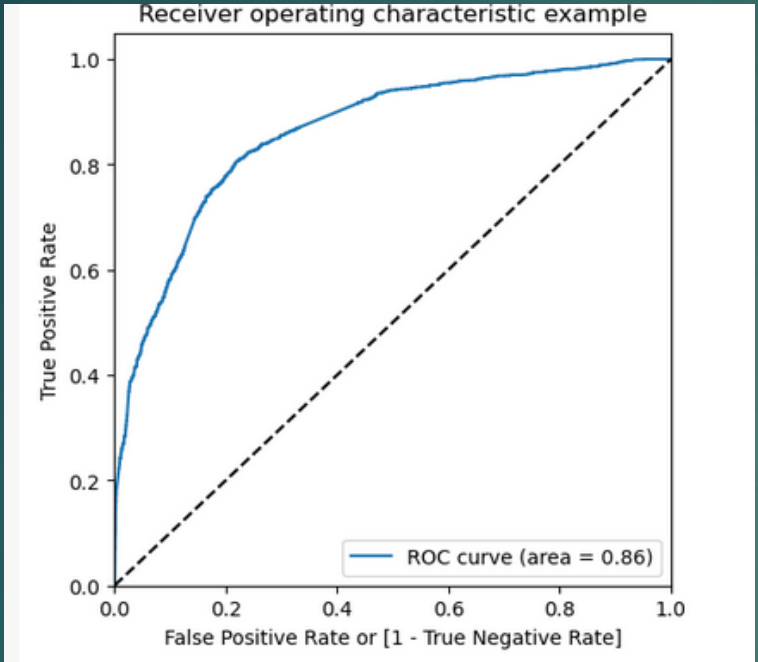
Data Conversion

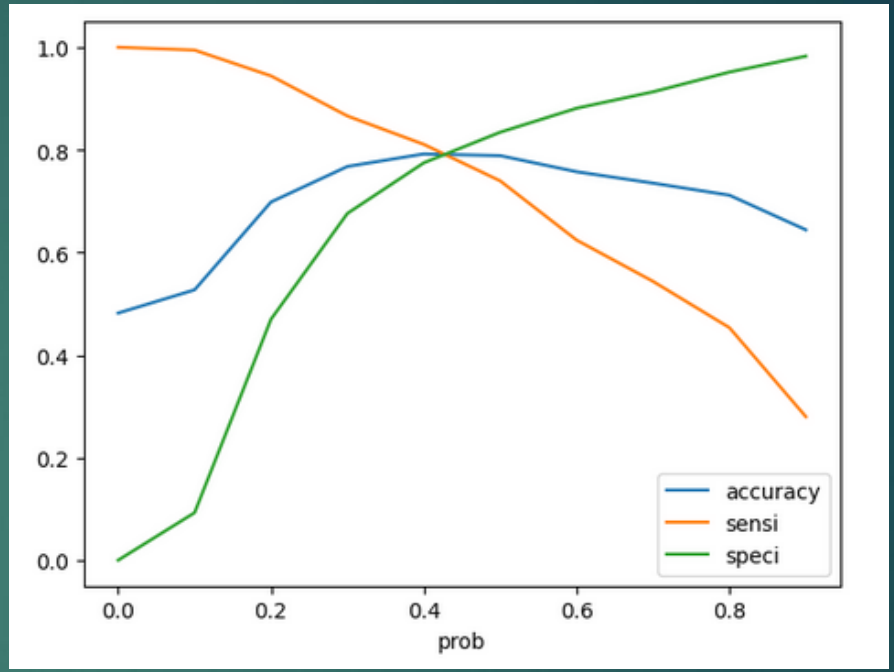
- Numerical variables have been normalized to ensure consistency in the analysis process.
- Dummy variables have been created for object type variables to facilitate statistical modeling.
- The dataset for analysis consists of a total of 8792 rows.
- A total of 43 columns will be considered for the analysis.
- These steps have been taken to ensure accuracy and efficiency in the analysis of the data.

Model Building

- Regression Data Splitting
- • Perform 70:30 train-test split.
- • Use RFE for feature selection.
- • Run RFE with 15 variables as output.
- • Build model by removing variables with p-value > 0.05 and vif > 5.
- • Predict test data set with 81% accuracy.

ROC Curve





- Optimal Cut Off Point Finding
- • Balanced sensitivity and specificity.
- • Graph shows optimal cut off at 0.35.

Conclusion

Potential Buyers' Factors in Website Visits

- Total time spent on website.
- Total number of visits.
- Lead source: Google, direct traffic, organic search.
- Last activity: SMS, Olark chat conversation.
- Lead origin: Lead add format.
- Current occupation: Working professional.
- X Education can thrive by influencing potential buyers to purchase courses.

THANK YOU