ENHANCING LEAD CONVERSION AT X EDUCATION

A PREDICTIVE MODELLING APPROACH

Sibani Panda

PROBLEM STATEMENT

Company Overview: X Education sells online courses to thr industry professionals.

Lead Generation:

- Professionals land on the website via marketing efforts on various platforms.
- Leads are generated through form submissions and past referrals.

Current Challenge:

- Lead conversion rate is only 30%.
- Out of 100 leads, only about 30 convert.

Objective:

- Improvisation of lead conversion rate to around 80%.
- Identification of 'Hot Leads' who are more likely to convert.

Solution Requirement:

- Development of a model to assign a lead score to each lead.
- Focusing on sales efforts on high-scoring leads to increase conversion efficiency.

Data Provided:

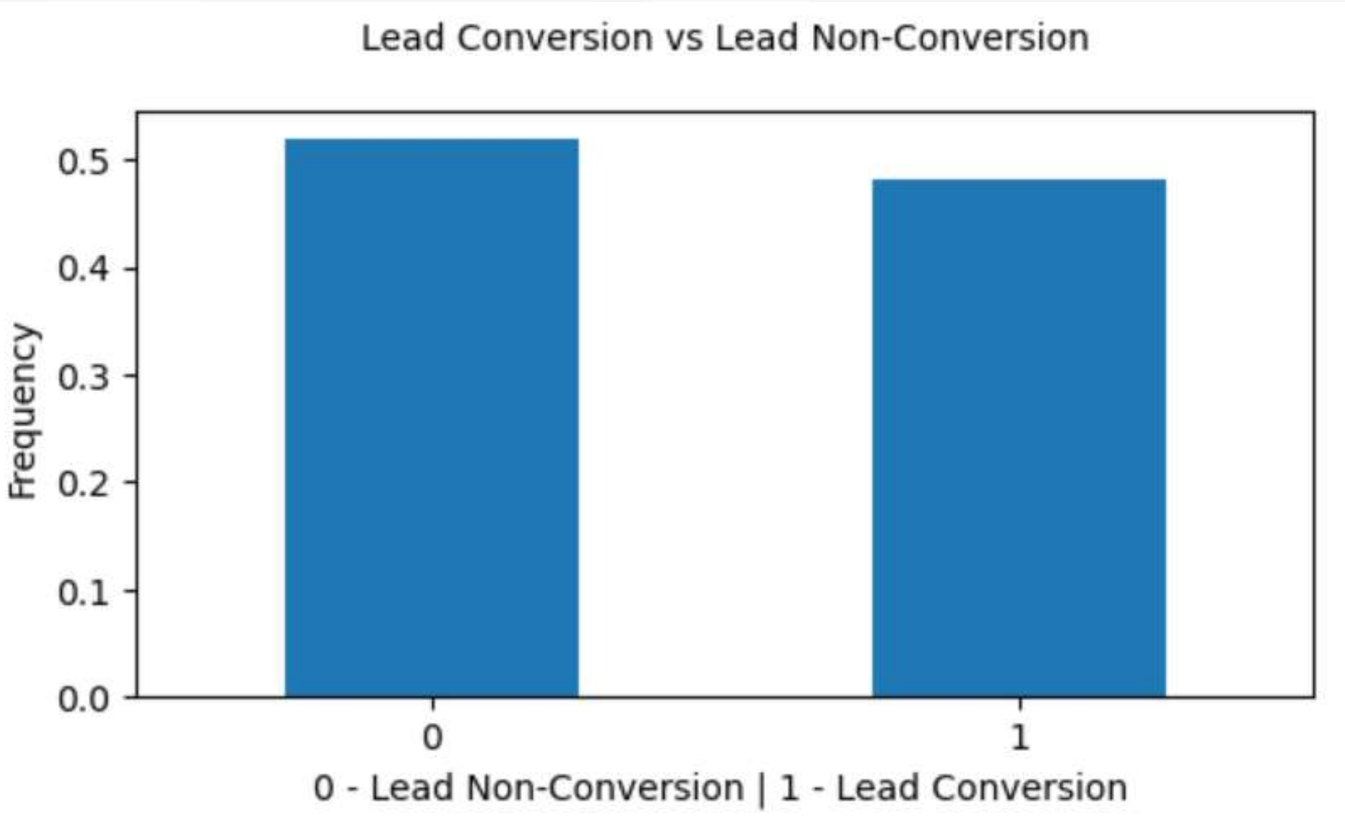
- Past leads dataset with 9000 data points & 37 attributes
- Includes customers' activity on the website and demographics.
- Target variable: 'Converted' (1 = converted, 0 = not converted).

PROPOSED SOLUTION

- Develop a logistic regression model to assign lead scores between 0 and 100.
- Determine an optimal cutoff score to identify high-potential leads.
- Sales team focuses on leads with scores above the cutoff.
- Higher scores indicate higher conversion likelihood (hot leads).
- Lower scores indicate lower conversion likelihood (cold leads).
- Prioritize contacting hot leads to improve conversion rates.
- Aim to significantly increase X Education's lead conversion rate from 30% to around 80%.
- Utilize data-driven insights to enhance lead management and sales efficiency.

OVERVIEW OF THE DATA

- Analyzed past leads data with around 9,240 records.
- Data includes details like where leads came from, time spent on the website, and their activities.
- Target information indicates if a lead converted to a customer (1) or not (0).
- Cleaned the data and removed unnecessary information.
- Treated outliers to ensure accurate analysis.
- Checked for data imbalance to ensure reliability.
- Prepared the data for building a predictive model.



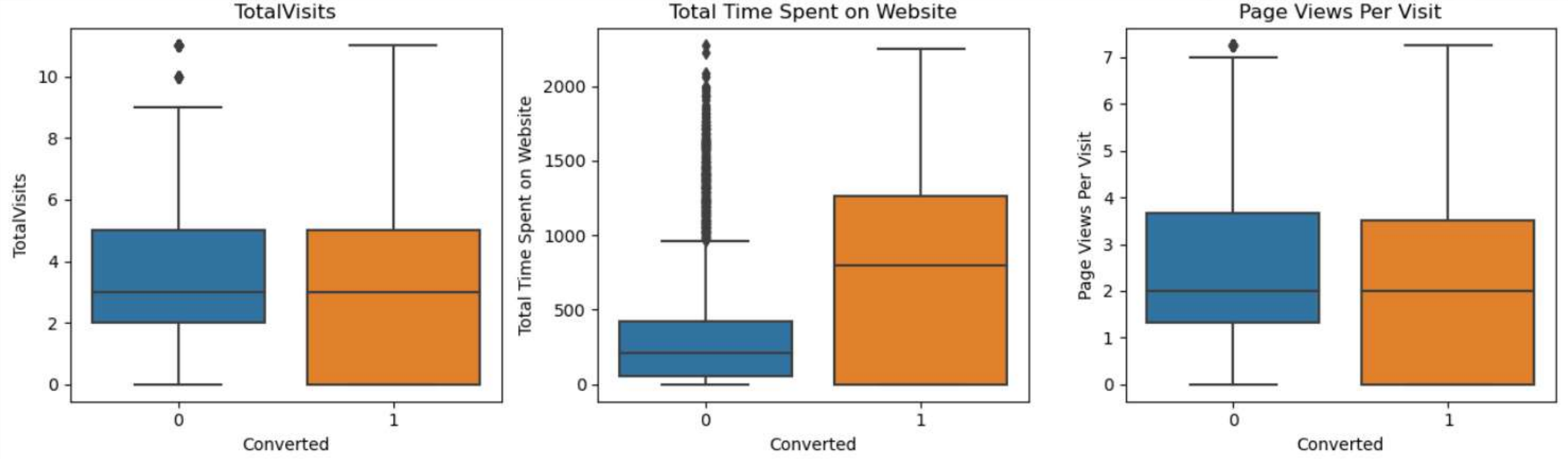
Data Imbalance Ratio :
1.08

For every 1.08 non-converted leads, there is 1 converted lead.

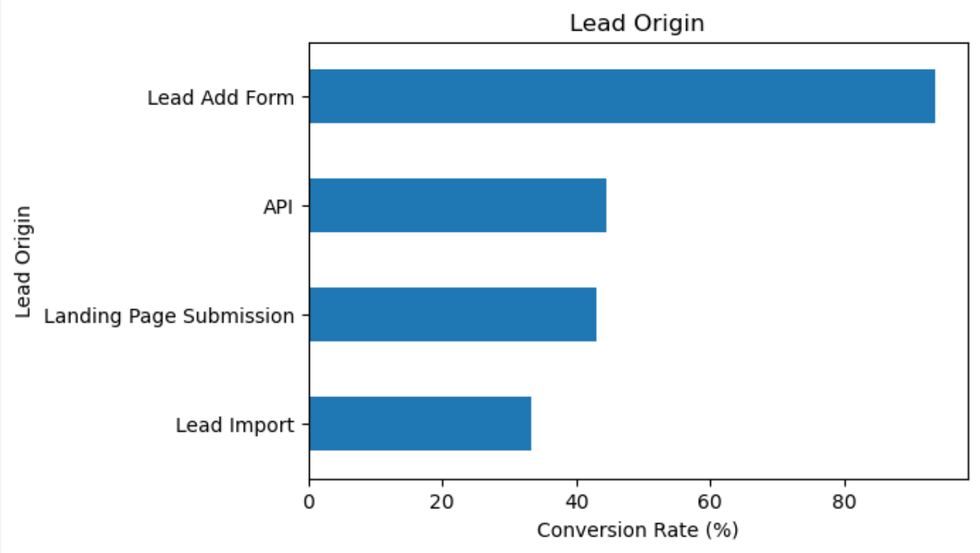
EXPLORATORY DATA ANALYSIS

IMPACT OF WEBSITE ENGAGEMENT METRICS ON LEAD CONVERSION RATE

Acquired customers have higher total visits, spend more time on the website. & have more page views per visit.

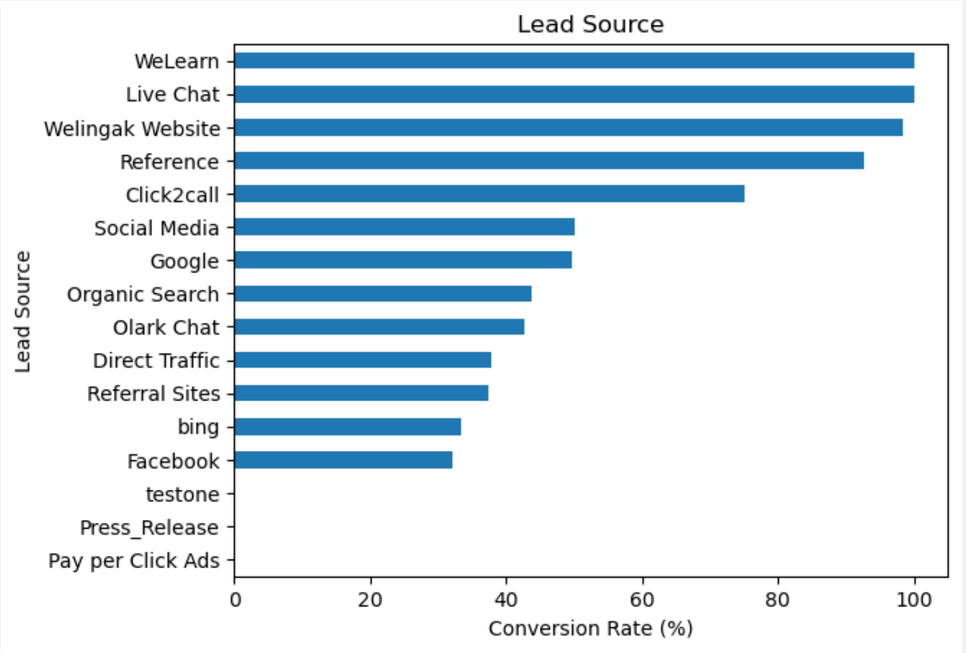


CONVERSION ANALYSIS BASED ON LEAD ORIGIN AND SOURCE



Most conversions originate from the direct 'Lead Add Form'.

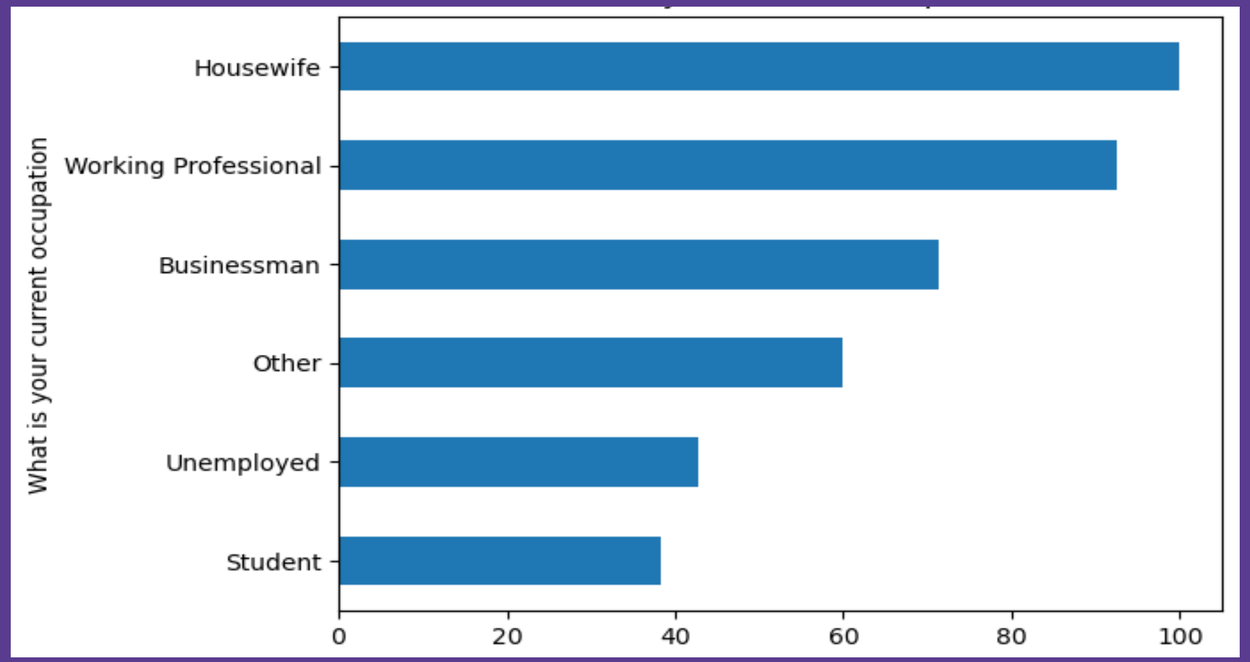
Highest conversions from sources such as Live Chat, WeLearn, and Welingak Website

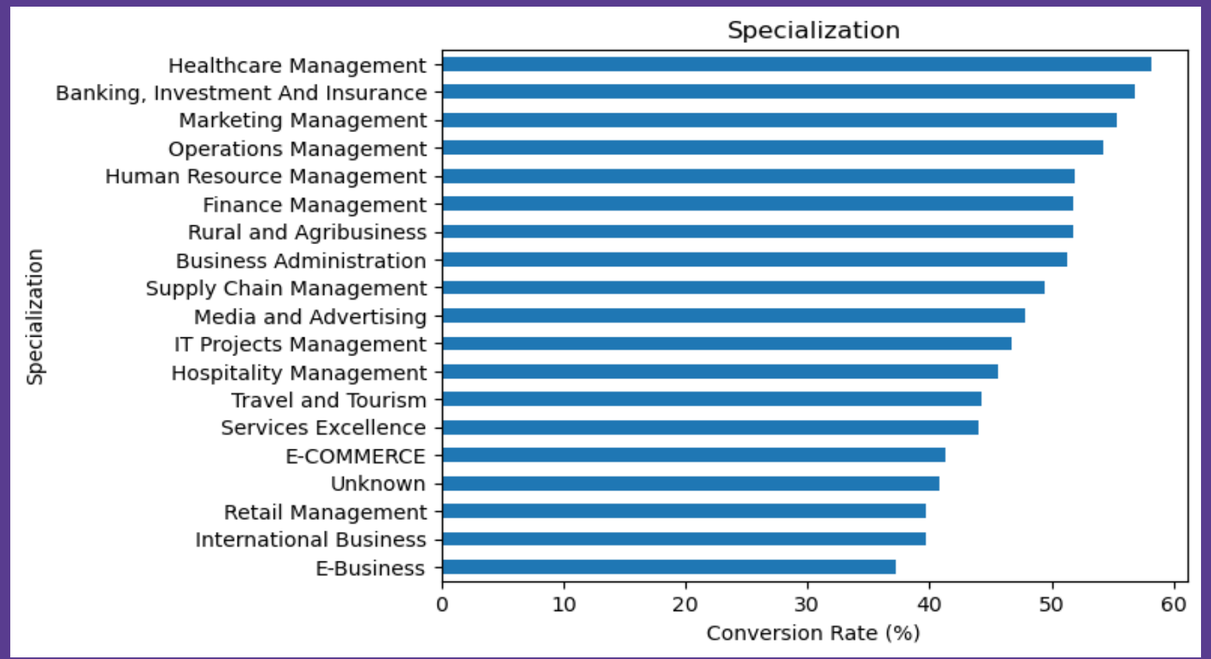


CONVERSION ANALYSIS BASED ON CUSTOMERS' DEMOGRAPHICS

Most leads from Healthcare, Banking, Marketing; fewest from E-Business.

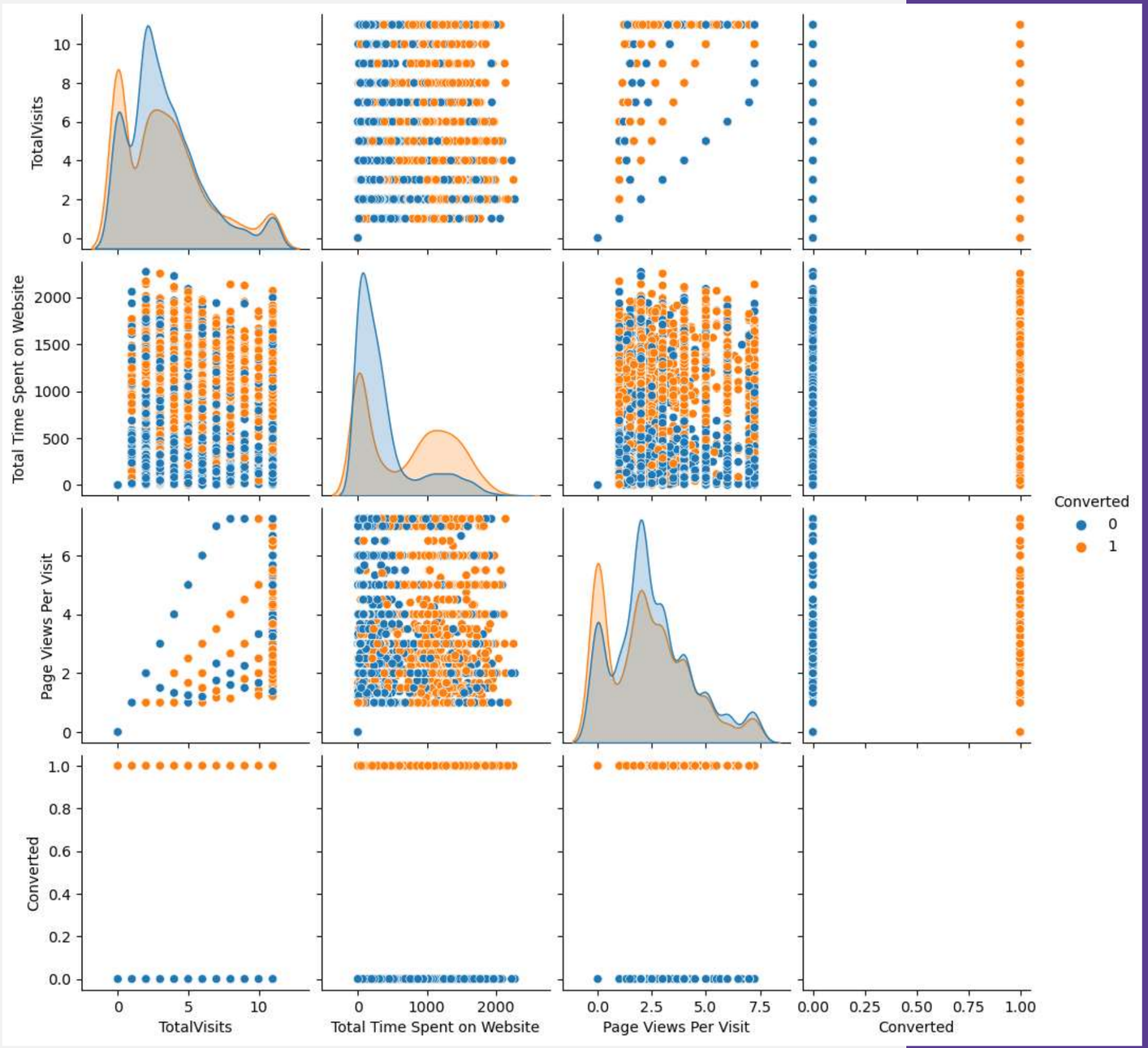
Highest conversions from Housewives, Working Professionals; lowest from Unemployed, Students.





CORRELATION ANALYSIS OF WEBSITE ENGAGEMENT METRICS AND CONVERSION RATES

- Longer website visits correlate with higher lead conversions.
- More page views per visit increase conversion likelihood.
- Total visits alone don't strongly predict conversions.
- Leads with higher engagement are more likely to convert.

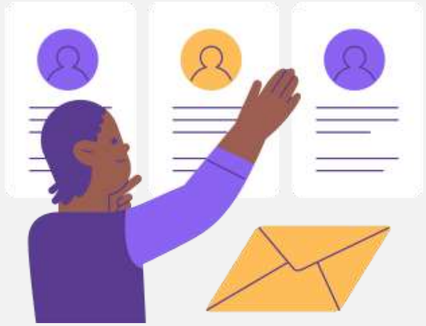


STREAMLINING LEAD ANALYSIS WITH KEY FEATURE IDENTIFICATION

Selected top 15 features impacting lead conversions using RFE. (Recursive Feature Elimination) Technique

Date has been split into training (70%) and testing sets (30%) for model learning and evaluation.

A predictive model will be developed using top 15 features; iteratively removing the less significant features using p-value (> 0.05) and VIF (> 5)



P-value: Measures feature's importance in predicting lead conversions.
VIF: Checks for redundant information among features to ensure reliability.

TOP 15 PREDICTORS/DRIVERS SELETED BY RFE

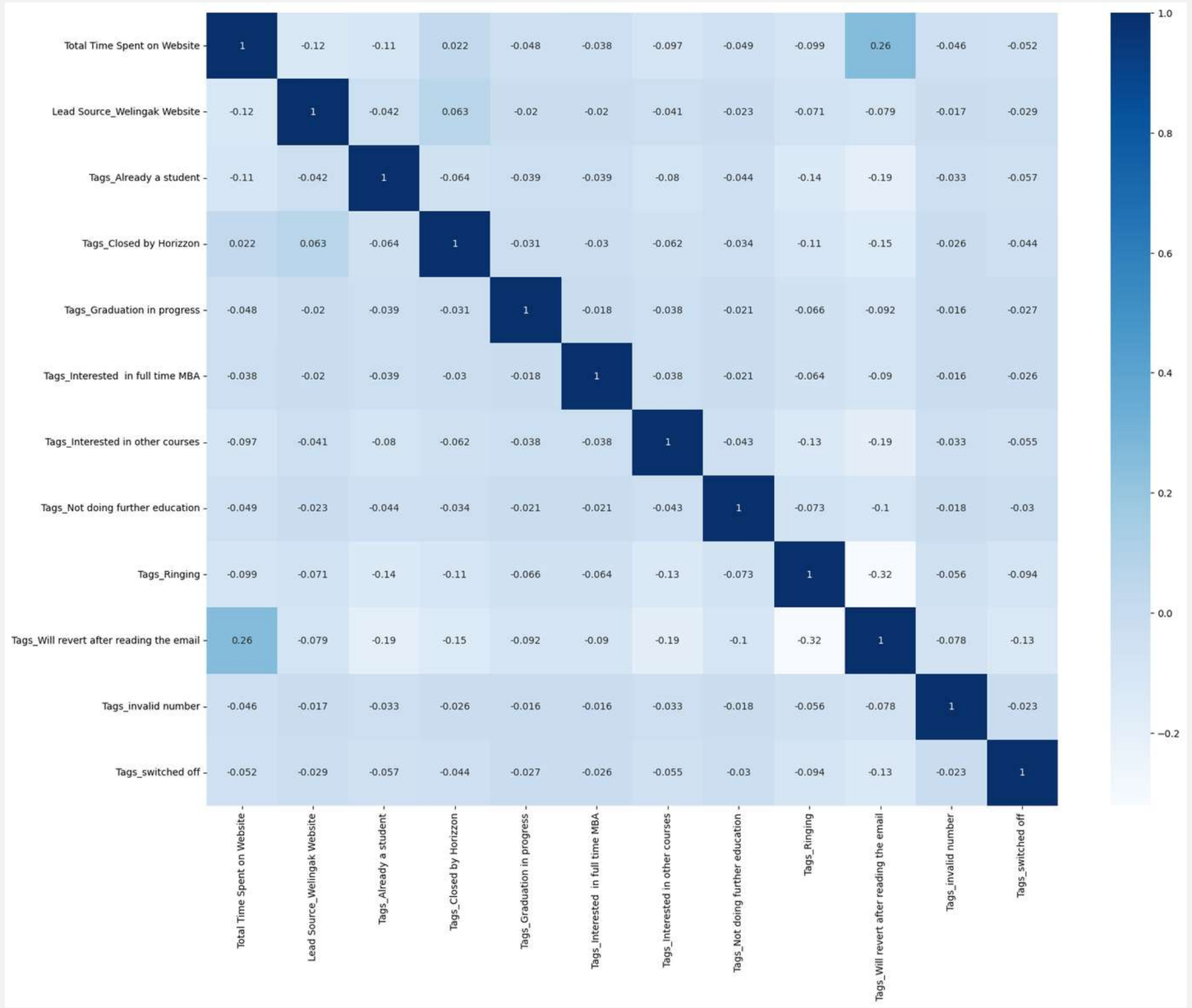
Total Time Spent on Website
Lead Source - Welingak Website
Tag - Closed by Horizzon (CRM Tool)
Tag - Diploma Holder (Not Eligible)
Tag - Graduation in Progress
Tag - Interested in Full-Time MBA
Tag - Interested in Other Courses
Tag - Not Doing Further Education
Tag - Ringing
Tag - Will Revert After Reading the Email:
Tag - Invalid Number
Tag - Number Not Provided
Tag - Will Revert After Reading the Email
Tag - Switched Off
Tag - Wrong Number Given

ANALYZING RELATIONSHIPS AMONG KEY LEAD CONVERSION FACTORS

- **Low Multicollinearity**: Independent variables show weak correlations with each other, indicating minimal multicollinearity. This is positive for model accuracy.

- **Key Relationship Identified**: A moderate positive correlation (0.26) between "Total Time Spent on Website" and "Tags_Will revert after reading the email" suggests that leads spending more time on the site are more likely to follow up.

- **Potential Redundancy**: A moderate negative correlation (-0.32) between the customers whose phone is ringing and customers who confirmed they will revert after reading the email" may indicate some overlap in the information provided by these variables.



STATISTICS OF THE FINAL MODEL

- **Significant Variables:** All model independent features have p-values of 0.00 indicating that the included variables are statistically significant.

- **Positive Co-efficient:** Key variables like "Total Time Spent on Website", "Lead Source", Tags - Closed by Horizzon have strong positive impacts, ensuring effective lead prediction.. A positive coefficient means that higher values for the variable increase the lead score, indicating a higher likelihood of conversion.

- **Negative Co-efficient:** Negative coefficients for specific tags highlight that customers with these factors are less likely to convert.

Variable	Coefficient	Std. Error	z-Value	p-Value
Intercept (const)	-0.3626	0.094	-3.868	0.000
Total Time Spent on Website	2.7472	0.255	10.754	0.000
Lead Source_Welingak Website	3.9572	0.723	5.476	0.000
Tags_Already a student	-5.3391	0.715	-7.469	0.000
Tags_Closed by Horizzon	4.4261	0.716	6.178	0.000
Tags_Graduation in progress	-2.9621	0.477	-6.206	0.000
Tags_Interested in full time MBA	-3.9152	0.726	-5.394	0.000
Tags_Interested in other courses	-3.9020	0.369	-10.562	0.000
Tags_Not doing further education	-4.8418	1.011	-4.791	0.000
Tags_Ringing	-3.9004	0.228	-17.083	0.000
Tags_Will revert after reading the email	3.1235	0.174	17.993	0.000
Tags_invalid number	-4.2650	1.020	-4.180	0.000
Tags_switched off	-4.6952	0.719	-6.534	0.000

MODEL EVALUATION ON THE TRAINING SET

Default Cut Off : 0.5

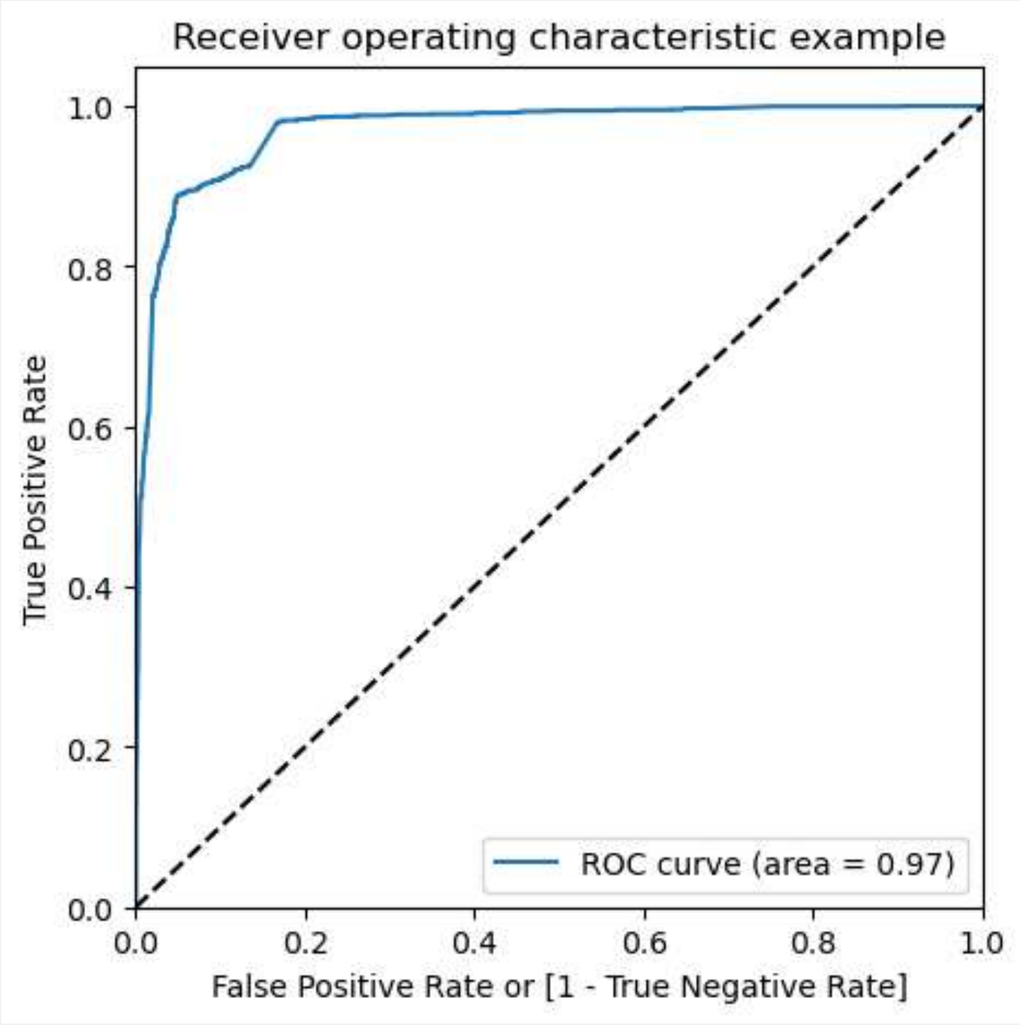
Overall Accuracy : 91%

Sensitivity : 90%

Specificity : 93%

Area under ROC curve : 0.97

It indicates model accuracy in distinguishing between classes. An area of 0.97 means excellent lead classification performance.



MODEL EVALUATION ON THE OPTIMAL CUT OFF POINT

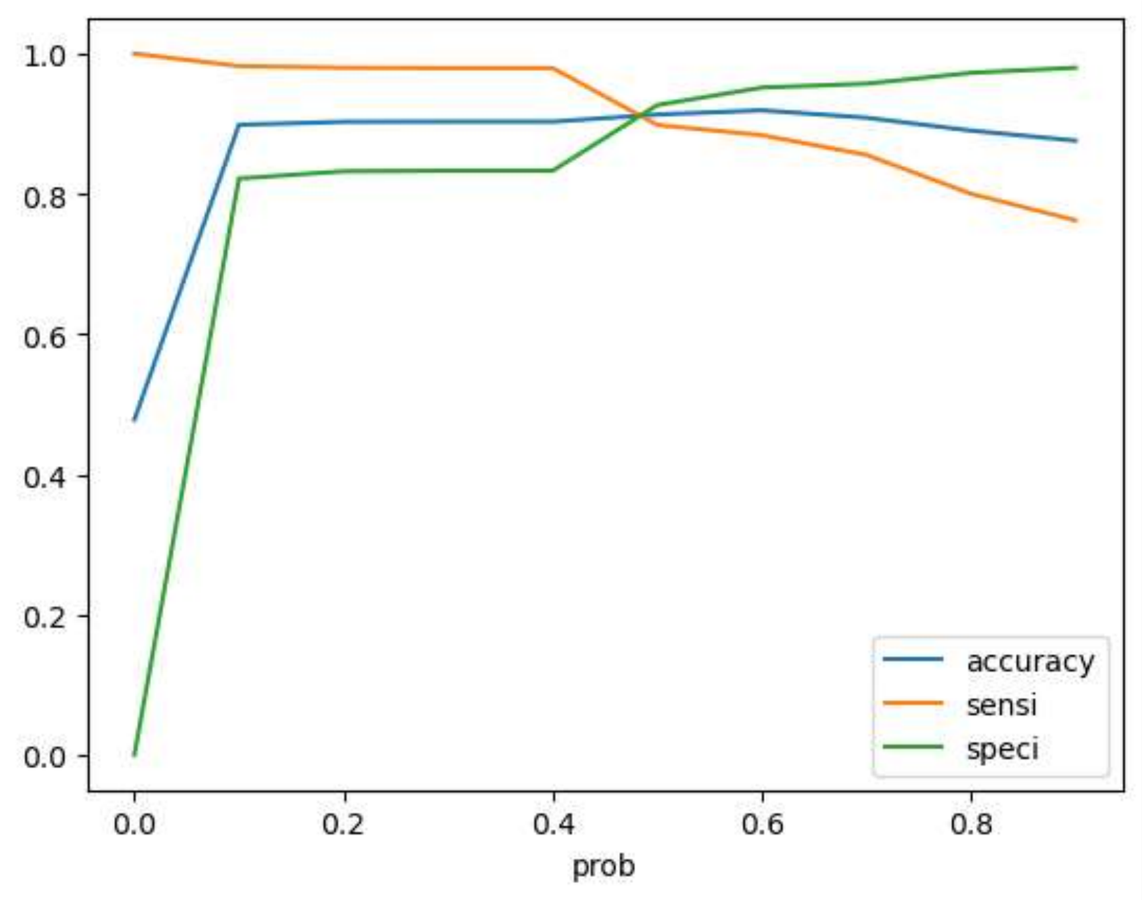
Optimal Cut Off
0.48

Overall Accuracy : 90.6%

Sensitivity : 90.7%

Specificity : 90.6%

Optimal cutoff balances accuracy, sensitivity, and specificity to maximize lead conversion prediction, ensuring effective targeting and minimizing errors.



PREDICTIONS ON THE TEST DATASET

	Prospect ID	Converted	Converted_prob	Lead_Score	final_Predicted
0	2828	1	0.973278	97	1
1	6562	0	0.014435	1	0
2	7468	0	0.027996	3	0
3	6337	0	0.017375	2	0
4	8566	1	0.973342	97	1

Using the test set, we predicted conversion class, conversion probability, and assigned lead scoring for each Prospect ID/Customer

Accuracy : 89.6%

model correctly classifies 89% of the test data.

Sensitivity : 90%

Identifies 90% of potential customers accurately.

Specificity : 90%

Correctly classifies 90% of non-potential customers

Precision : 87.8%

Leads marked as potential are genuinely potential

Recall : 91.4%

Identifies 91% of actual potential customers correctly

STRATEGY RECOMMENDATION

1. Consider Leads from CRM Tool Horizzon: Leads closed by Horizzon tool has higher conversion probability

2. Prioritize Website Leads: Target leads coming from the Welingak Website for effective engagement.

3. Target Preferred Communication: Engage leads who prefer email communication for better response rates.

4. Focus on Engaged Leads: Prioritize those who spent more time on the website, as they show higher conversion potential.

5. Follow Up on Email Responses: Prioritize customers who have confirmed they will respond after reading emails for timely follow-up.

6. Segment by Occupation: Focus on housewives, working professionals, and leads closed by Horizon for higher conversion chances.


Thank
you