Enhancing Lead Conversion at X Education

Comprehensive Analysis and Strategic Recommendations

PROBLEM STATEMENT

Objective: Increase lead conversion rate for X Education

Current Situation: Average conversion rate is around 30%.

Goal: Achieve a conversion rate of 80% by identifying high-potential leads.

Approach Overview

- **Data Preparation**: Cleaning and pre-processing the dataset to handle missing values.

- **Exploratory Data Analysis (EDA)**: Understanding data distributions and relationships.

- **Model Building**: Developing a logistic regression model and fine-tuning it.

- **Lead Scoring**: Assigning scores to leads based on predicted conversion probabilities.

Data Preparation

•**Handling Missing Values**: Dropped columns with >40

% missing data; imputed others with mean/mode.

```
In [11]: # Check for missing values and calculate the percentage
         missing_percentage = data.isnull().sum() * 100 / len(data)
         print(missing_percentage)

         Prospect ID                                        0.000000
         Lead Number                                        0.000000
         Lead Origin                                        0.000000
         Lead Source                                        0.389610
         Do Not Email                                       0.000000
         Do Not Call                                        0.000000
         Converted                                          0.000000
         TotalVisits                                        1.482684
         Total Time Spent on Website                        0.000000
         Page Views Per Visit                               1.482684
         Last Activity                                      1.114719
         Country                                            26.634199
         Specialization                                     36.580087
         How did you hear about X Education                 78.463203
         What is your current occupation                    29.112554
         What matters most to you in choosing a course      29.318182
         Search                                             0.000000
         Magazine                                           0.000000
         Newspaper Article                                  0.000000
         X Education Forums                                 0.000000
         Newspaper                                          0.000000
         Digital Advertisement                              0.000000
         Through Recommendations                            0.000000
         Receive More Updates About Our Courses             0.000000
         Tags                                               36.287879
         Lead Quality                                       51.590909
         Update me on Supply Chain Content                  0.000000
         Get updates on DM Content                          0.000000
         Lead Profile                                       74.188312
         City                                               39.707792
         Asymmetrique Activity Index                        45.649351
         Asymmetrique Profile Index                         45.649351
         Asymmetrique Activity Score                        45.649351
         Asymmetrique Profile Score                         45.649351
         I agree to pay the amount through cheque           0.000000
         A free copy of Mastering The Interview             0.000000
         Last Notable Activity                              0.000000
         dtype: float64
```

```
# Verify no remaining missing values
print(data.isnull().sum())

Prospect ID                                        0
Lead Number                                        0
Lead Origin                                        0
Lead Source                                        0
Do Not Email                                       0
Do Not Call                                        0
Converted                                          0
TotalVisits                                        0
Total Time Spent on Website                        0
Page Views Per Visit                               0
Last Activity                                      0
Country                                            0
Specialization                                     0
What is your current occupation                    0
What matters most to you in choosing a course      0
Search                                             0
Magazine                                           0
Newspaper Article                                  0
X Education Forums                                 0
Newspaper                                          0
Digital Advertisement                              0
Through Recommendations                            0
Receive More Updates About Our Courses             0
Tags                                               0
Update me on Supply Chain Content                  0
Get updates on DM Content                          0
City                                               0
I agree to pay the amount through cheque           0
A free copy of Mastering The Interview             0
Last Notable Activity                              0
dtype: int64
```

Before

After

Exploratory Data Analysis (EDA)

- **Target Distribution**: Most leads are not converted; focus on characteristics of converted leads.
- **Key Insights**:
  - High engagement (Total Visits, Total Time Spent on Website) correlates with conversions.
  - Significant categorical variables include Lead Source and Last Activity.
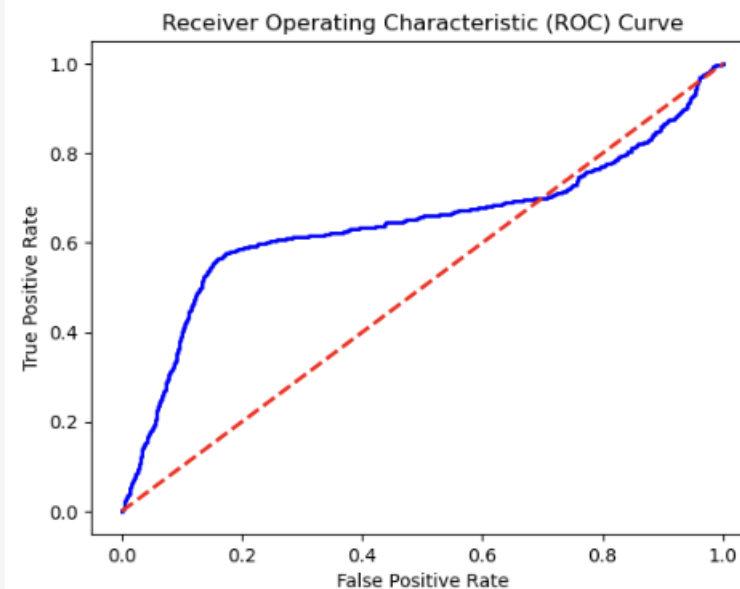

Distribution of Converted Leads

Correlation Matrix
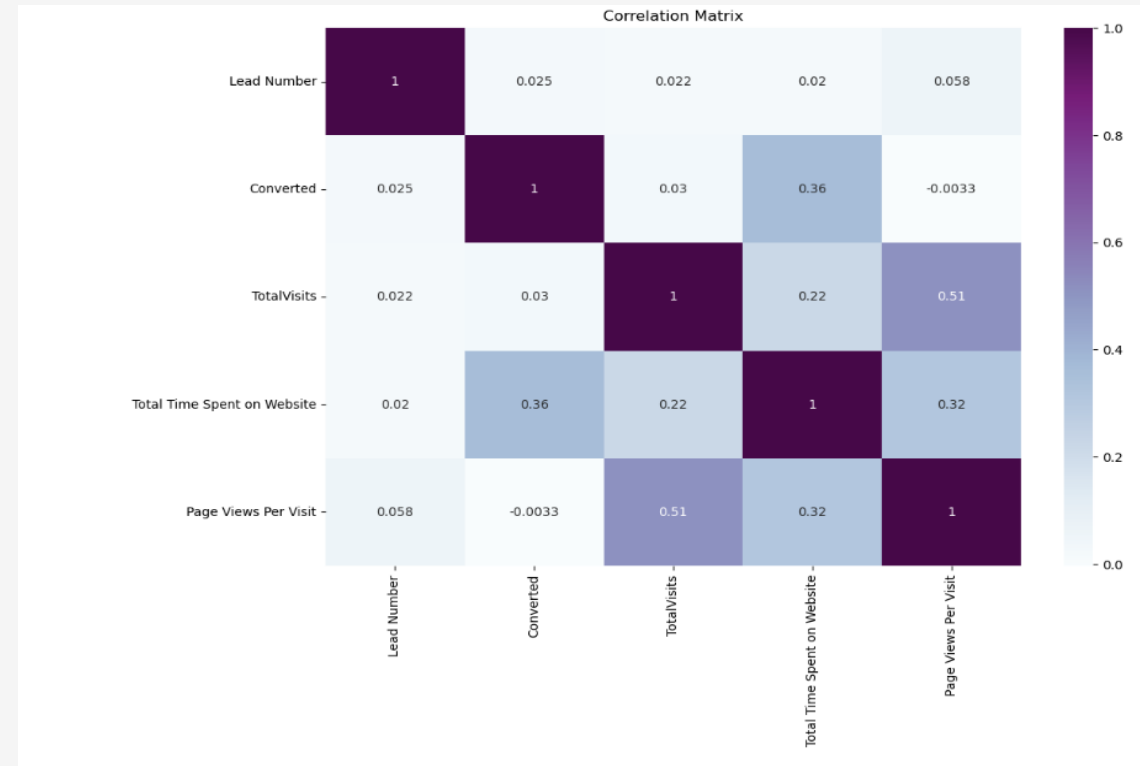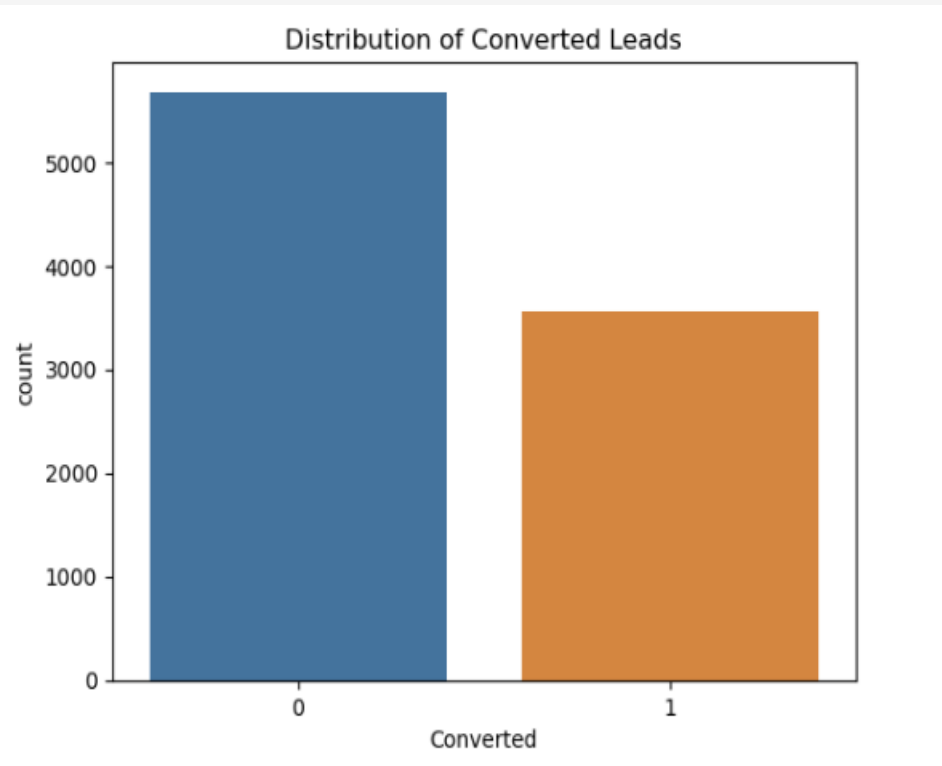
Model Building

•**Initial Model**: Logistic Regression **Performance Metrics**:

- Accuracy: 73%

- ROC AUC Score: 0.64 (room for improvement).


```
[[1473  222]
 [ 540  537]]
              precision    recall  f1-score   support

           0       0.73      0.87      0.79      1695
           1       0.71      0.50      0.58      1077

    accuracy                           0.73      2772
   macro avg       0.72      0.68      0.69      2772
weighted avg       0.72      0.73      0.71      2772
```
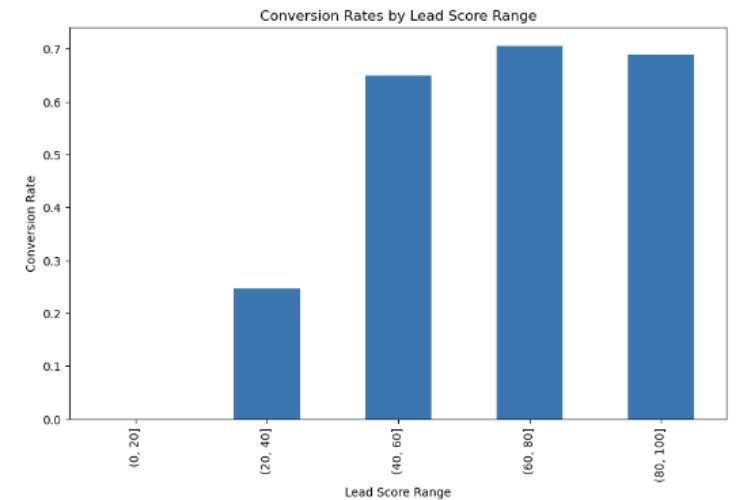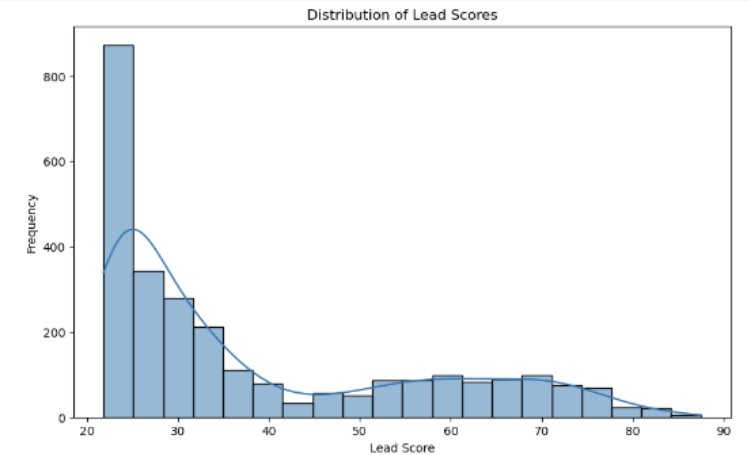
Lead Scoring

•**Method**: Assign lead scores based on model's predicted probabilities.

•**Analysis**:

- High-potential leads identified with scores ≥ 80.

- Conversion rates increase with higher lead scores.

	Lead Score	Converted
4608	30.281355	1
7935	54.321595	0
4043	23.186374	0
7821	33.539416	0
856	27.383699	0

Distribution of Lead Scores

Conversion Rates by Lead Score Range

Key Findings and Recommendations

- **Top Features**:

  - Total Time Spent on Website

  - Total Visits

  - Last Activity

- **Top Categorical Variables**:

  - Lead Source

  - Last Activity

  - Specialization

- **Strategic Recommendations**:

  - Aggressive conversion strategies during intern phases.

  - Minimize outreach during periods when targets are met by focusing on high-potential leads.

Business Impact

- **Efficiency**: Focus on high-potential leads to boost conversion rates.

- **Scalability**: Model adapts to new data, continuously improving lead conversion strategies.

- **Next Steps**:

  - Implementing recommendations.

  - Continuous model monitoring and fine-tuning.