Purpose: Improve lead conversion rate for X Education

Data Preprocessing

- Handled missing values: Dropped columns with >45% missing data and filled missing values in 'Tags', 'City', 'Lead Source', 'Specialization'
- Removed irrelevant features: 'Prospect ID', 'Lead Number', 'What matters most to you in choosing a course', 'Lead Profile', 'What is your current occupation', 'Country', 'How did you hear about X Education'
- Created dummy variables for categorical features
- Handled outliers in 'TotalVisits' and 'Page Views Per Visit'

Exploratory Data Analysis (EDA)

- Most leads from India (before dropping 'Country')
- 'Google' is the primary lead source
- 'Landing Page Submission' and 'API' are major lead origins
- 'Email Opened' and 'SMS Sent' are common last activities

Model Development

- Logistic Regression model
- Used Recursive Feature Elimination (RFE) to select top 15 features
- Applied StandardScaler to numeric columns
- Used statsmodels for detailed model statistics

Key Predictors of Lead Conversion

Total Time Spent on Website, Total Visits, Last Activity: SMS Sent, Lead Origin: Landing Page Submission, Lead Source: Direct Traffic, Specialization: Others

Model Performance

- Accuracy - 90.2
- Sensitivity -81.18
- Specificity - 95.6
- ROC -.96

Business Recommendations

- Focus on increasing website engagement (time spent and visits)
- Optimize SMS marketing campaigns
- Improve landing page for lead submissions
- Enhance direct traffic channels
- Tailor approach for leads from various specializations