Summary: This analysis was conducted for X Education with the objective of increasing enrollment of industry professionals in their courses.

Key Steps:

1. **<u>Importing the datasets</u>** : The provided data yielded valuable insights into the website's visitor behavior, including their visit patterns, duration, referral sources, and conversion rates.

2. **<u>Cleaning the dataset</u>** : The data, mostly clean, required addressing null values and replacing the uninformative 'select' option with null. Some nulls were labeled 'not provided' to retain data, but these were later removed during dummy variable creation. To streamline location data, entries were categorized into 'India,' 'Outside India,' and 'not provided.'
   Please find exact steps done for cleaning the dataset:
   1. checking for the no. of unique values in the categorical columns
   2. dropping the columns with highly skewed data
   3. finding the percentage of null values present in the columns
   4. dropping the columns with null values which is higher than 40 percent of the total values
   5. standardizing the values
   6. Mapping Binary categorical variables
   7. spotting for the outliers
   8. capping and flooring the outliers

3. **<u>Exploratory Data Analysis (EDA</u>**): A swift EDA highlighted irrelevant elements in categorical variables, while numeric values appeared sound with no outliers detected.
   Please find exact steps done for EDA:
   1. checking for the data imbalance
   2. Univariate Analysis for Categorical Variables
   3. Bivariate analysis of the categorical columns
   4. Bivariate Analysis for Numerical Variables

4. **<u>Data preparation:</u>**
   Please find exact steps done for Data preparation:
   1. creating the dummy variable

2. splitting the data : Train-Test Split: The dataset was divided into 70% training and 30% testing subsets.
3. Scaling of the data

5. **Model Building**: RFE identified the top 15 relevant variables, and the remaining variables were manually removed based on VIF and p-value criteria (VIF < 5 and p-value < 0.05 were retained).

Please find exact steps done in model building:

1. We will Build Logistic Regression Model for predicting categorical variable : 6 models are created
2. Feature Selection Using RFE (Coarse tuning)
3. Manual fine-tuning using p-values and VIFs

6. **Model Evaluation**: A confusion matrix was constructed, and the optimal cutoff value (determined via ROC curve) resulted in approximately 80% accuracy, sensitivity, and specificity.

1. Confusion Matrix
2. Accuracy
3. Sensitivity and Specificity
4. Threshold determination using ROC & Finding Optimal cutoff point
5. Precision and Recall

7. **Prediction**: 80% for the metrics we are getting with the sensitivity-specificity cut-off threshold of 0.359. So, we will go with sensitivity-specificity view for our Optimal cut-off for final predictions.

8. **Precision-Recall**: As we can see in above metrics when we used precision-recall threshold cut-off of 0.42 the values in True Positive Rate, Sensitivity, Recall have dropped to around 76%, but we need it close to 80% as the Business Objective.

Key Findings: Analysis revealed the most influential factors for potential buyers:

Top 3 features that contributing positively to predicting hot leads in the model are:

- Lead Source_Welingak Website
- What is your current occupation_Working Professional
- Lead Origin_Lead Add Form

Top 3 features that contributing negatively to predicting hot leads in the model are:

- Lead Origin_Landing Page Submission
- Specialization_not provided
- What is your current occupation_not provided

Increasing Lead Conversion Rates:

- Feature Focus: Prioritize features with positive coefficients for targeted marketing strategies.
- Quality Leads: Develop strategies to attract high-quality leads from top-performing sources.
- Tailored Messaging: Engage working professionals with personalized communication.
- Optimize Channels: Analyze lead engagement impact and optimize communication channels accordingly.
- Budget Allocation: Allocate more budget for advertising on the Welingak website.
- Incentives for Referrals: Offer incentives or discounts for successful referrals that convert to leads.
- Target Working Professionals: Aggressively target working professionals due to their higher conversion rates and potentially better financial situations.
- Enhance Website Appeal: Improve the website's attractiveness and engagement to encourage longer visits and increase the chances of course purchases.

Areas for Improvement:

- Current Occupation: Address the issue of missing data for "current occupation."
- Specialization Analysis: Investigate negative coefficients related to specialization and prompt users to provide this information.
- Landing Page Optimization: Review the landing page submission process for potential areas of improvement.

Considering these key factors, X Education can enhance its success by effectively targeting potential buyers and encouraging them to enroll in their courses.