Summary

A dataset containing 9,240 rows and 37 potential features was analyzed to derive insights and develop a predictive machine learning (ML) model. The analysis followed a systematic approach, and the key findings are summarized below:

1**. Data Review and Preparation**

- Data Quality: Approximately 30% of the dataset was excluded due to missing or invalid entries, such as placeholder values ("Select"). The remaining 70% was used for analysis.
- Recommendation: To enhance future datasets, critical fields should be made mandatory during data collection to ensure data completeness and accuracy.

2**. Analytical Approach**

The analysis was conducted in three stages:

- Data Cleaning: Addressed inconsistencies and prepared the dataset for modeling.
- Exploratory Data Analysis (EDA): Identified trends and relationships within the data.
- Model Development:
    - Dummy variables were created.
    - Data was split into training and testing sets.
    - Features were scaled for consistency.
    - A logistic regression model was developed using techniques such as Recursive Feature Elimination (RFE), Generalized Linear Models (GLM), Variable Inflation Factor (VIF), and p-value validation.
    - The model was evaluated and prediction metrics verified.

3. **Key Insights**

- Lead Distribution:
    - The majority of leads were from India, with Mumbai as the most represented city.
    - Leads were evenly distributed across various specializations, indicating broad interest in the course offerings.
- Lead Behavior:
    - Approximately 70% of leads expressed interest in the courses to advance their careers.
    - Website metrics, such as page visits and time spent, were critical predictors of lead conversion.
    - Communication preferences were dominated by phone calls and SMS, highlighting the need to explore alternative engagement methods.

4. **Model Findings**

- A logistic regression model was developed to predict lead conversion.
- The model determined an optimal probability threshold of 0.42 for predicting conversions.

5. **Strategic Recommendations**

- Improve data collection processes by making critical fields mandatory to avoid unusable entries.
- Maintain and enhance the website's user experience (UX) and content to boost engagement.
- Unemployed leads formed a significant portion of the dataset but had a conversion rate of only 42%. Revisiting course pricing and commitment levels could enhance their appeal.