Summary

1. A dataset containing over 9,000 rows was thoroughly reviewed and analyzed to extract actionable insights and support the development of a machine learning (ML) model.

2. High-Level Process
- **Data Cleaning and Preparation:** Ensured data quality and readiness for analysis.
- **Exploratory Data Analysis (EDA):** Identified patterns and relationships in the data.
- **Model Building:**
  - Creation of dummy variables.
  - Splitting the data into training and testing subsets.
  - Feature scaling to normalize the data.
  - Model generation using techniques like Recursive Feature Elimination (RFE), Generalized Linear Models (GLM), and Variable Inflation Factor (VIF), with p-value analysis for validation.
  - Model evaluation and prediction metric verification.

3. Post data cleaning, approximately 69% of the dataset was determined to be suitable for building the ML model.

4. Key insights include:
- A significant proportion of leads originated from India, with Mumbai as the primary city.
- Around 70% of leads expressed interest in the course to enhance their career prospects.
- Many fields in the dataset contained invalid placeholder values (e.g., "Select"), which rendered them unusable. It is recommended to make these fields mandatory during data collection to improve data quality and insights.
- The leads were evenly distributed across various specializations, indicating the course catalog appeals to diverse professional backgrounds.
- A large share of leads came from unemployed individuals, followed by working professionals. Introducing more competitive pricing may enhance conversion rates among unemployed leads.

5. A **Logistic Regression Model** was developed to predict lead conversion into paying customers. The model determined an optimal probability threshold of 0.43 for lead conversion prediction.

6. Behavioral insights:
- Website metrics, including page visits and time spent, were strong predictors of lead conversion. Enhancing the website's content and user experience (UX) is crucial for sustained success.
- Communication modes such as phone calls and SMS remain dominant in engaging leads.

7. Strategic Recommendation: Maintaining an informative website, improving mandatory data field completion, and offering more attractive pricing for targeted demographics will likely enhance lead conversions and overall effectiveness of the course offerings.