Presentation Summary

Problem Statement: The primary objective of this analysis is to identify the key factors that influence student enrollment (lead conversion) in an online professional training course. By understanding these drivers, we can refine our marketing strategies to effectively target prospective students, enhance their engagement, and ultimately improve the course's enrollment rates.

Analysis Approach:

1. Data Cleaning: We handled missing values, outliers, and categorical variables when cleaning and organizing the data. We also looked into correlations and data distributions.
2. Feature Engineering: Using pre-existing features, such as visit/conversion ratios and engagement ratios, we developed new ones.
3. Model Construction and Selection: Two logistic regression models were constructed, one utilizing RFE feature selection and the other with all features. We contrasted how well they performed on test data that was not visible.
4. Model Interpretation: To determine the relative significance and direction of each feature's influence on conversion, we examined the final model's coefficients.

Results:

We successfully identified several key factors that significantly impact lead conversion:

1. Website Engagement: Users who demonstrated higher engagement—characterized by frequent visits, longer time spent on the site, and a higher number of page views per visit—were more likely to convert.
2. Specialization Interest: Individuals interested in management specializations had notably higher conversion rates compared to those exploring other areas.
3. Occupation: Users who were unemployed or still students showed a greater likelihood of enrolling in the course, possibly due to their motivation to enhance skills and improve job prospects.
4. Geographical Location: Conversion rates were found to be higher among users located in Mumbai and the wider state of Maharashtra, indicating a strong regional demand for the course.
5. The model using RFE-based feature selection slightly outperformed the full-feature model. This improvement in performance suggests that focusing on the most relevant features and excluding less important ones not only enhances the model’s accuracy but also improves its interpretability, making the results easier to act upon.

Visualizations:

We used a variety of visual tools to aid in understanding and presenting the data:

1. Histograms and boxplots were utilized to visualize the distribution of continuous variables and detect the presence of outliers.
2. Heatmaps helped us examine correlations between different features, identifying relationships that could guide further analysis.
3. Feature importance charts were employed to clearly display the relative influence of each feature on the likelihood of lead conversion, helping stakeholders grasp which factors warrant the most attention.

Conclusion:

This analysis effectively identified the primary factors that influence lead conversion for the online professional training course. By focusing on increasing website engagement, targeting specific user groups (such as unemployed individuals, students, and those interested in management specializations), and intensifying marketing efforts in key regions like Mumbai and Maharashtra, the course provider can enhance enrollment rates and drive substantial business growth. These insights can inform future marketing campaigns and guide strategic decisions to further optimize customer acquisition efforts.