SUMMARY

In this case study, we aimed to analyze lead conversion data and provide strategic recommendations to improve the conversion rate. The process began with importing necessary libraries, loading the dataset, and understanding the problem clearly. After setting up the environment, we proceeded with data processing, which involved handling missing values and outliers to maintain data quality. Exploratory Data Analysis (EDA) followed, using univariate and bivariate methods to explore patterns and relationships within the data, offering key insights into lead behavior.

In the data preprocessing phase, we created dummy variables to convert categorical features into numerical formats. We also removed unnecessary columns and duplicate entries to ensure clean data for modeling. We then defined X as the predictor variables and y as the target variable, representing whether a lead converted or not. The dataset was split into training and testing sets using a train-test split to allow for evaluating the model's performance on unseen data. Features were scaled using StandardScaler to ensure that all variables were on the same scale, which is essential for machine learning models like logistic regression.

For feature selection, Recursive Feature Elimination (RFE) was used to identify the most significant predictors for lead conversion. We validated the selected features by checking their p-values, retaining those with p-values less than 0.05. Additionally, Variance Inflation Factors (VIF) were checked to ensure multicollinearity was not an issue, with acceptable VIF values below 5. After building the logistic regression model, we evaluated its performance using accuracy, recall, and other relevant metrics. This allowed us to identify the variables that had the most significant impact on lead conversion.

Key insights and recommendations were derived from this analysis. First, we recommended prioritizing leads from certain sources, such as "Welingak Websites" and "Olark Chat," as these leads showed a higher likelihood of conversion. Second, targeting working professionals, who often seek upskilling opportunities, was suggested, as they are more likely to convert. We also advised focusing on leads with higher engagement, such as those spending more time on the company’s website, as these individuals demonstrated higher interest in the company’s offerings. Additionally, following up with leads whose last activity involved SMS or phone conversations was recommended, as they displayed increased potential for conversion.

We also identified leads to deprioritize. Leads whose primary course decision factors were classified as "Other" showed a lower likelihood of converting, as did leads from "Landing Page Submissions" and those with the "Others" specialization. Lastly, we recommended respecting the preferences of leads who selected "Do Not Email," as they explicitly expressed a desire to limit communication.

In conclusion, by focusing on high-potential leads and avoiding low-conversion segments, the company can optimize its sales strategy and significantly improve its lead conversion rates. Prioritizing leads from effective sources, engaging professionals, and focusing on those who have shown interest will help drive better results while respecting the preferences of less-

engaged leads. Improved engagement strategies, smarter targeting, and tailored communication are essential for boosting conversion success.