

## **Executive Summary**

The goal of the Lead Scoring Case Study is to help X education company with recommendations on how to improve their lead conversion rate.

The approach followed was to begin by understanding the data and checking for any formatting issues where Select category was found representing null values. After replacing them with NaN, columns having more than 40% null values were removed. Next missing values were handled in categorical variables having greater than 5% missing values by creating another category as Other. For the rest of the categorical columns, missing values were replaced with mode. Missing values in numeric variables were replaced with median due to presence of outliers. Outliers were replaced using upper and lower limit of IQR. Next, columns with high data bias were removed. Since conversion rate of data is 38%, separate data frames were created for EDA. In order to prepare data for model building, categories were reduced for dummy variables creation. Columns having high multi-correlation based on correlation matrix were removed. After splitting in test and train data frames, numeric variables were standardized using standard scaler. RFE was used to select top 15 significant variables. Columns having high p-value were reduced one by one. ViF was checked for variables to ensure least multi-collinearity. Cut-off point was decided as 0.3 by plotting accuracy, sensitivity and specificity. Accuracy, confusion matrix, precision score and recall score were used as metrics for model evaluation. Lead score values were derived on the test based on the conversion probability.

Leads that spent more than 600 seconds on the website are more likely to convert. Will revert after reading the email and Closed by Horizon are common tags for converted leads. SMS Sent Activity has resulted in good conversion rate. Last Activity, Last Notable Activity, Tags and City have different categories contributing to converted and not converted leads, thus, they can be good factors. Landing page submission with total visits 4 has highest converted leads whereas Landing page submission with total visits 2 has highest non converted leads.

The model results indicate that even if a lead is Closed by Horizon, Lost to Others, in touch with EINS or Lost to EINS, there is still a chance for them to be converted. Providing incentives and offers to such leads can result in conversion. Hence, they can be considered in the target audience. Also, for the leads whose reply is awaited i.e. in confusion whether part time or DLP, Still Thinking or Will revert after reading the email, a follow up with doubts resolution or alumni session can be helpful in increasing the chances of conversion. Since a good amount of leads imported from other sources are cold, time investment on such leads should be minimal.