Lead Scoring Case Study Summary:-

We did the analysis for X education to look into the leads they have generated and help them figure out the most efficient ways to improve their sales.

1. **Cleaning Data** :- We first clean the data. We look at the Null values, we remove the columns for the variables who had less then 2% null values. We converted 'Select" to NaaN. We dropped all the variables if the null values were more than 40%. For rest of the variables with null values we imputed them with most important (highest in percentage) values in the variable.
2. **EDA:-** In EDA, we did Univariate and Bivariate Analysis. Multiple variables were found who are not relevant for the analysis hence we dropped those columns. Outliers were taken care of and the numeric values looked good.

3. **Preparing Data for Modelling**:- we created dummy variables for 'Lead Origin', 'Lead Source', 'Last Activity', 'Specialization', 'What is your current occupation', 'City' and 'Last Notable Activity' and then dropped the columns for whom dummy variables were created

4. **Splitting the data into Train and Test:-** The data was split in 70% and 30% for train and test data

5. **Model building and evaluation:-** First we do feature selection using RFE and then started building models. We built 9 models and based on p-values and VIF values, we kept on removing variables. By the end of making model 9, P-value of all the variables has come to 0 and VIF values are also very low. Also, we are left only with 12 variables by now A confusion matrix was also made. Later on the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around (on Train data) **81%**, **81.7%** and **80.6%**.

6. **Prediction:-** It was done on the test data frame with an optimum cutoff of **0.34.** and accuracy, sensitivity and specificity of **80.4%, 80.4% and 80.5%.**

7. **Precision – Recall:-** This was also used to recheck and a cutoff of 0.41 was found with Precision **79.5%** and Recall of **70.6%**

It was found that we can achieve the lead conversion rate of around 80% using this model. Also, it was found basis this model we should contact the leads with lead score of >85. There are around 368 leads who have a really high chance to get converted into an actual sale.

Insights:-

The organization should focus more on generating leads through Welingak Website & Refe rences and they have a much higher chance to convert into a Sale

People who are working Professionals have a much higher chance to get converted in to a sales hence we must focus more on convincing working professional

The company should make calls to the leads who spent "more time on the websites" as th ese are more likely to get converted.

The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.

The company should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.

They should also focus more when the lead source was **Google, Organic Search or Direct t raffic to website.**