SUMMARY : LEAD SCORING CASE STUDY

This study is done for X Education to encourage more industry professionals to join their course. The company requires a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The basic data provided with information about customer behaviour, demographics, actions, sources etc.

Approach to Model Building :

1. **Data Cleaning and Preparation** : After replacing the missing & 'Select' values with NaN, the columns with more than 40% of missing values were excluded from the dataset as they become insignificant in our analysis. The columns with less than 5% of missing values were replaced with mode of their respective columns. Further, for columns with missing values ranging between 6% to 39% were analysed individually. The Outliers with 5% and 95% quantiles were excluded for numerical variables. Further, many columns with unique values, single values and extreme data imbalance were excluded.

2. **Exploratory Data Analysis** : Univariate, Bivariate and Multivariate data analysis was conducted and each variable analyzed with respect to Converted column. Important column identified are Lead Source, Lead Origin, Last Activity, Specialization, Tags and numerical variables like Total Visit, Total Time Spent on Website, Time Sent Per Page.

3. **Data Preparation for Model Building** : Dummy Variables created, Feature Scaling done, Training & Testing datasets created

4. **Model Building** : RFE is done to maintain top 15 variables. Gradually the variables were dropped considering their P value less than 0.01 and VIF less than 5.

5. **Model Evaluation** : ROC Curve : 0.97

	TRAINING DATA	TESTING DATA
ACCURACY	92.01%	92.39%
SENSITIVITY	90.92%	92.69%
SPECIFICITY	92.68%	92.19%
PRECISION	88.44%	88.56%
RECALL	90.92%	92.69%

6. **Important Variables considered to be relevant for lead conversion** :

   • Tags (Closed by Horizzon, Lost to ENIS, Will Revert after reading the Email)
   • Lead Source (Welingak Website, Reference)

- Last Activity (SMS sent)
- Total Time Spent on Website
- Total Visits

7. **Recommendations** : Depending the important variables the target customers must be identified and these customers must be attended with personalized experience such as the information must be kept handy and elaborate about the desired products. Monitor each of these lead carefully so that you can tailor the information you sent to them. A carefully drafter plan which caters the needs to each of these leads will go a long way to improve probability of lead conversion. Hold open discussion forums and informative session for resolution of enquiries and FAQs.