

Lead Score Case Study Summary Report: -

Problem Statement:

X Education Online company who sells online courses to industry professionals are looking for **increase** in their lead **conversion rate** currently from **30% to ~80%**.

Goals of Assignment: -

1. **Identify most important factors** contributing to higher conversion rate.
2. Identify **focus areas** to increase probability of conversion.

Steps followed in Assignment:

1. **Initial Analysis** of Data.
2. **Data Cleaning** to handle missing records, duplicate records and outliers.
3. **EDA and Visualisation** of data using Bivariate and Multivariate Analysis by plotting graphs.
4. **Preparation** of Data including Splitting of data to Train-Test dataset and standardizing numerical column using StandardScaler method.
5. Created correlation matrix of train dataset and plotted heatmap to identify highest collinearity of features.
6. **Eliminated multicollinear features** using combination of automated method (RFE) and manual method (VIF) values.
7. **Iteratively built model** by removing features that are not significant to dependent variable (Converted) after each step.
8. **Finalised best fit model** with **Accuracy** score of **94%**, **Sensitivity** score of **90%** and **Specificity** score of **96%**.
9. Obtained **optimal cut-off** value from **Sensitivity-Specificity trade off plot** which is **0.32**.

Findings from Assignment:

1. A few features had the term "**Select**," which stands for "**Missing values**" in the context of the default choice that shows when a customer do not provide any details. We **changed Select labels into Missing values** for further processing.
2. '**Google**' accounted for **32% of the company's leads**, while 'Direct Traffic' came in second with 27%, and 'Olark Chat' accounted for 19% of the leads.
3. The highest percentage of **leads' last activity, 39%**, was "**Email Opened**." **SMSs sent by leads** contributed to **30%** as their last activity.
4. **Maximum Leads (60%) are Not Working** that is they are either Student or Unemployed.
5. **Converted Leads** tend to **spend more time on Website** and they view more pages and **visit more websites** as compared to non-converted customers.

Below are the top three variables that contribute most towards probability of lead getting converted: -

1. Total time spent on Website.
2. Page Views Per Visit.
3. Lead Source.

Recommendation for X Education company to increase probability of Lead Conversion:

1. Leads who learned about company via source 'Welingak Website' have more probability to become customers and hence company should **focus on broadcasting feedback and ratings from previous clients on Welingak Website** that will have an impact on increase of conversion rate.
2. Leads who have 'SMS sent' as their last activity recorded have more chance in getting converted. Company can **increase publishing attractive marketing campaigns via SMS** for these categories of leads so that they have high chance of getting converted to customers.
3. Working Professional Leads have higher chance of buying online courses. Company can focus on **providing offers or rewards for Working Professionals Leads** so that they can buy and avail those offers which will ultimately increase probability of conversion.