

SUMMARY

Our Goal :

As the initial goal was to achieve a 50% increase in conversion rate from the prevailing conversion rate of 30% and assign a lead score between 0 to 100 to each potential lead. Also identifying various factors which help in driving more productive conclusions. We created a methodical approach as below to achieve our targets:

1. Data Understanding :
 - a. Import data and analyse on all the features descriptions and interpretations to better judge on the problem statement.
2. Exploratory Data Analysis
 - a. After getting a gist of the dataset, imputing missing values and data cleaning would help us create more robust model.
 - b. Univariate and Multivariate Analysis to understand the relationships/patterns between the independent and our dependent variables, also to deal with outliers to prevent our model from being skewed.
3. Dummy Variable Creation :
 - a. This step involves converting all significant categorical variables into an encoded form so that model is able to interpret the classes of each variable.
4. Numerical Variable Scaling :
 - a. All numerical variables had different scales of measurement while data sourcing hence it creates unequal weightage for each variable at the time of modelling. To bring equal contribution we scale all numerical variables to improve performance.
5. Test and Control Split :
 - a. To prevent overfitting and better evaluation we create the train and test split.
6. Logistic Regression Modelling :
 - a. Automated feature elimination, manual feature elimination, model evaluation on the training set, ROC AUC analysis, finding the optimal cut-off to improve model accuracy, precision recall curve analysis, and testing the model with the optimal cut-off on test data.
7. Conclusion and Learnings
 - a. The problem to solve here was we had to increase the lead conversion rate from 30% to 80% by analysing the potential leads in a more productive manner. As per the initial analysis Mumbai came out to be the top priority for targeting potential leads. Maybe as the company will grow we will get more geographical targets but as of now we should target all the existing leads which are majority from Mumbai.
 - b. Though it's not a clear majority, Finance management, HR, Marketing are the most worked professions which brings more potential leads and maybe marketing in these specific domains could be fruitful.
 - c. The ratio of leads to learner is the best for working professional, although a healthy amount of unemployed did get converted but non conversion is still quite high. Therefore if company have limited resources to acquire a learner maybe they should target working professionals.

- d. The best mode of targeting leads can be through SMS/Email since the most of the conversions came through these two channels as Last activity. While Calling had a negative impact on conversion since it have a negative variable importance.
- e. One clear cut correlation we can observe is with the amount of time spent on the website. It's what the company should monitor since that could really help boost the conversion rate and identify prospects.
- f. Also we can observe the most contributing variables after the RFE were **Tags, Last Activity, Lead Source and Lead Origin** in a positive way.
- g. Among the most important features, best tags with high variable importance were
 - i. Closed by Horizzon
 - ii. Lost to EINS
 - iii. Will revert after reading the email.

Overall, we can say we should target our first set of leads with the best correlated variables that are, to 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which will filter the very best prospects of potential leads. Although our RFE eliminated these variables but our EDA showed healthy level of correlation between them and our target variable. Post which, we can target further in terms of location, domains in which they are interested in/working currently, check if they are working or unemployed, send emails and SMS to generate the curiosity of exploring our courses. Tags, Last Activity, Lead Source, Lead Origin as some of the major important variable the company should consider before approaching a lead. Marketing channels should place more relevant course descriptions in offices, corporate locations, etc to attract working professionals since they have a better conversion. More hoardings and advertisements can be targeted in Mumbai and related cities to boost reach. The sales team should maintain and update Tags, Last Activity, Lead Source, Lead Origin since if they were to use our models these variables would be key in identifying a potential lead.