LEAD SCORE CASE STUDY:

PROBLEM STATEMENT:

1. Company X advertises the online courses to theircustomers and aims at maximizing their profit.
2. The company has taken a detailed report of their customers and analyzed the trend which encompassesthe lead conversion proportion.
3. The customers are targeted by the company through several platforms like Google, Email , Phonefollow-ups etc, any customer who shows interest in the programs offered and ready to fill theform and accept the process of enrolment is considered as 'Leads' in business terms these are targeted customers.
4. The company's lead conversion rate is very poori.,e it has only 30% of lead conversion rate and this results in considerable loss to the company as the Return on Investment is hampered here
5. As the company head, we are asked to come up with a solution to identify the reason behind this and build a model which predicts the leads appropriately and in addition to that also we are asked to provide recommendations to increase the lead conversion rate.
6. The company has many features which drive the problem statement and a target variable 'Converted'. Based on the interactions of these variables, the lead score should be assigned for each customer between 0and 100 and based on the recommendations of the lead score, the company decides whether to target thespecific customer or not.

PROCEDURES FOLLOWED:

1. We identified the important variables of interest which impact the lead conversion rates directly and presented them briefly in the presentation.
2. Identified Categorical and Numerical variables of interest, also removed certain variables which do not have any values for the problem.
3. Handled outliers in the numerical data and removed the data whose percentile of occurrence is >99%.
4. Handled categorical variables and performed binary encoding for two-level variables and dummy encoding to multi-level variables.

5. Analyzed the correlation between the numerical variables and removed based on correlation matrix check to avoid multi-collinearity problem.
6. Built a logistic regression model and chosen the optimal cut-off point through ROC method.
7. Obtained the confusion metrics and inspected the metric measures like Sensitivity, Specificity, Recall, Precision, and Accuracy.
8. Have provided proper recommendations to the market about the dos and don'ts based on the results obtained.

LEARNING GATHERED:

1. We understood the importance of EDA , before building a model never mind how sophisticated it is , EDA is the most important part before model building.
2. Able to appreciate the importance of visualization and inferences made which paves the way not only to understand the hidden pattern in data but also able to correlate its significance towards decision making.
3. Learnt nuances about the model-building part and understood the significance of fine-tuning the cut-off point in the logistic regression to suit business needs.
4. Learnt to inspect the confusion metrics and applied the concepts to choose the metric carefully based on the problem statement.
5. Importantly we learnt how the business-related problem needs to be tackled and questions to raise to business to better understand the problem (in PPT provided in recommendation part)
6. Have learnt how to give business-related recommendations to the problem proposed and the importance of business presentation to present our view in layman terms to better suit the non-technical stakeholders.

General Recommendations Proposed to the Business:(Dos)

1. We can recommend based on the lead score generated , in the model the hot leads are given the score > 35 and those with less significant given the score <35%
2. We can also target the Lead Source channels like 'Olark Chart' , 'Google'
3. We could infer that the management type of courses are having higher lead conversion rates.

4. People with occupation as Working-Profession can be targeted and those with currently ‘unemployed’ also show greater conversion rates because they plan to be industry ready.
5. People who visit the website often generally show higher conversion rates (it is also categorized under the lead score) .

General Recommendations Proposed to theBusiness:(Don’ts)

1. Any customer whose lead score is less than 30% can be neglected as they are not categorized under hot leads and it is not worth in spending the resources on them.
2. Students generally show less conversion rates might be because of they are hesitant to bear the course fee, so it is good to avoid the students and concentrate with other professional folks.
3. The conversion rate in India is better compared to others, so let us not focus other countries until we see a satisfactory result in India.
4. We can reduce advertising in newspaper, Educational Forums as it does not add any value.
5. We analyzed that most of the customers do not like to be called / contacted through mobiles / instead it could be effective if the communication happens through email.
6. The customers don’t show any interest towards supply chain so we can temporarily stop advertising and let us review about the course contents and plan for improvement.