**Problem Statement:**

Employees in new-age insurance company perform multiple outbound efforts to sell term insurance to the customers. Telephonic marketing campaigns still remain one of the most effective ways to reach out to people however they incur a lot of cost. Hence, it is important to identify the customers that are most likely to convert beforehand so that they can be specifically targeted via call.

It is expected to build a model which will help the organisation by predicting the potential customers.

**Summary:**

**-Model Used**

Logistic Regression

**-Why Logistic Regression**

The given dataset is binary classification problem. It is a linearly separable dataset hence Logistic regression can be one of the best fit model.

**-Steps**

-Importing Dependencies

-Handling the Dataset

-EDA

-Elimination of Outliers

-Encoding

-Oversampling

-Model Development

**-Complication**

-Due to the imbalance in the dataset the output was skewed and the model was performing with low efficiency.

**-Optimization**

-To fix the imbalance and skewed output, I used the SMUTE Oversampling technique.

-Desired Sample Size used is 3.5 Times of "Yes" in the dataset.

**-Result**

-Before Oversampling the model accuracy was 68% and precision/recall/F1-Score valued between 68-71.

-SMUTE increased the Accuracy to 84% and precision/recall/F1-Score to 84.