Summary Report
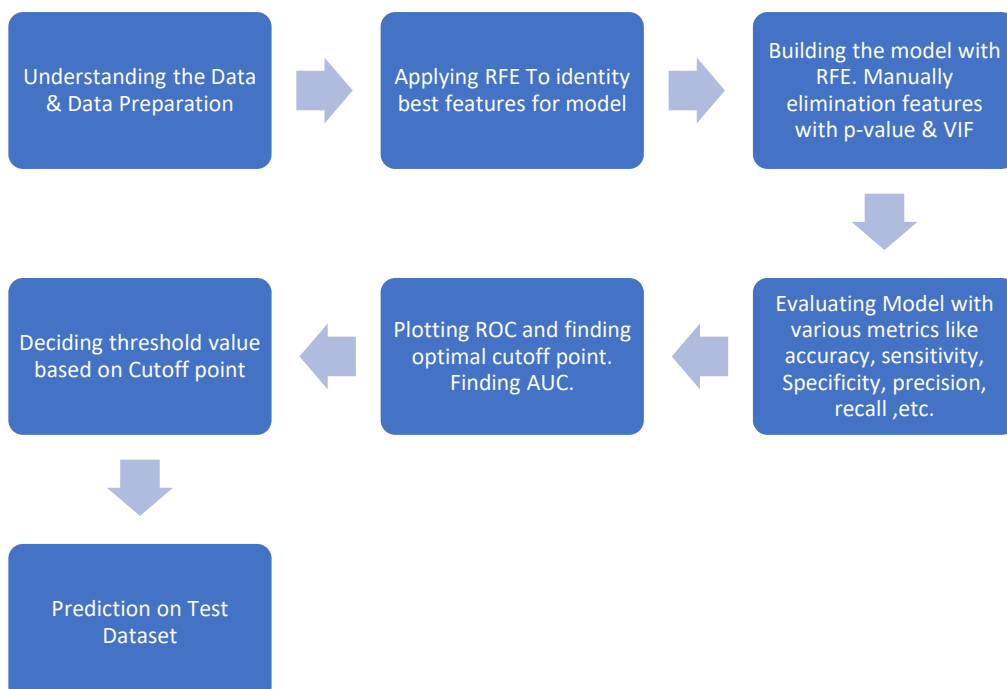
Following process is followed during the whole Case study.

1. Importing Necessary Libraries
2. Data preparation
   a. Data Reading
   b. Data Inspection
   c. Data Cleaning
3. Exploratory Data Analysis
   a. Univariate Analysis
   b. Bivariate Analysis
4. Model
   a. Feature Scaling
   b. Model Building
   c. Feature Selection using RFE and using VIF
   d. Prediction on Train Dataset
5. Model Evaluation
   a. Checking Accuracy
   b. Sensitivity
   c. Specificity
   d. FPR
   e. Precision
   f. Recall, etc.
6. Plotting ROC Curve and finding AUC and finding optimal cutoff point
   a. Assigning lead score
   b. Evaluating model again
7. Prediction on Test dataset

```
┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│ Understanding the   │ ──▶  │ Applying RFE To     │ ──▶  │ Building the model  │
│ Data & Data         │      │ identity best       │      │ with RFE. Manually  │
│ Preparation         │      │ features for model  │      │ elimination features│
│                     │      │                     │      │ with p-value & VIF  │
└─────────────────────┘      └─────────────────────┘      └─────────────────────┘
                                                                     │
                                                                     ▼
┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│ Deciding threshold  │ ◀──  │ Plotting ROC and    │ ◀──  │ Evaluating Model    │
│ value based on      │      │ finding optimal     │      │ with various metrics│
│ Cutoff point        │      │ cutoff point.       │      │ like accuracy,      │
│                     │      │ Finding AUC.        │      │ sensitivity,        │
│                     │      │                     │      │ Specificity,        │
│                     │      │                     │      │ precision,          │
│                     │      │                     │      │ recall ,etc.        │
└─────────────────────┘      └─────────────────────┘      └─────────────────────┘
          │
          ▼
┌─────────────────────┐
│ Prediction on Test  │
│ Dataset             │
└─────────────────────┘
```

Findings

After the final model building and evaluation, the prediction on test dataset was done. And we could get the following insights from it: -

- All variables have p-value < 0.05
- All features have VIF less than 5
- They final features selected hardy have any multicollinearity among them. That is evident from the heatmap.
- The probability threshold of 0.2 gives an overall accuracy on Test Dataset of about 0.9074 that is almost 91%
- AUC is around 0.95 that means the model is good
- Now we can say that for our particular model, as one increases, the other decreases and vice versa. Different values of the sensitivity and specificity can be achieved for the same model by changing the Conversion Probability cutoff threshold value.
- When the probability thresholds are very low, the sensitivity is very high and specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.
- High sensitivity implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the Conversion likelihood, i.e., it will misclassify some non-Conversion cases as Conversions.
- So, in order to increase the conversion rate less probability threshold should be used and vice-versa.