**Summary:**

1. **Import and perform Data Cleaning:**

Imported and performed the below operation

* Removed ‘Select’ values from Categorical columns, since the prospect did not choose any option while filling
* Dropped columns with more than 45% Nan values
* Replaced missing values with appropriate values for each individual columns
* Removed the rows which has less % of missing values

1. **EDA**

Performed EDA on few columns to see how the Conversion rate varies with different values of Categorical features

1. **Data Preparation:**

* Created Dummy variables for Categorical features
* Split the Test and Train data with 0.7 and 0.3 split size
* Did features scaling using Standard scaler

1. **Model Building**

* Used RFE to do feature selection
* Brought down the features to 15
* Build the model with these features using Stats model GLM method
* Removed features with high *p* value like *Last Activity\_Had a Phone Conversation* and *Occupation\_Housewife* and brought down the required features to 13.
* Verified the collinearity by calculating VIF and found that all the selected features are non collinear.

1. **Prediction**

* Performed prediction and generated Probabilities using the above model
* Initially took 0.5 as cutoff for conversions and generated Conversions
* Generated Confusion Matrix
* Calculated Accuracy, Specificity and Sensitivity. Generated ROC curve. The Auc has come out to be 0.89
* Plotted Accuracy, Specificity and Sensitivity with different cutoff’s and plotted to see the optimal cutoff.
* It is identified that 0.4 is optimal.

1. **Final Model**

* Used 0.4 as cutoff and regenerated the conversions
* Generated Confusion matrix
* Final values of metrics
  + Accuracy: 0.81
  + Sensitivity/Recall: 0.76
  + Specificity: 0.84
  + AUC(Area under the curve): 0.89
  + Precision: 0.75

1. **Applying on Test Data**

* Scaled the test data based on trained scaler
* Applied the model on Test data in required feature columns
* Generated Confusion matrix and calculated different metrics
* Final metrics on Test data are
  + Accuracy: 0.81
  + Sensitivity/Recall: 0.76
  + Specificity: 0.85
  + Precision: 0.74