1. Refined insights: What are the most meaningful insights from the data relevant to the problem?

- LogReg juxtaposed features that positively affect Defaulter status: DEROG, JOB_Self, NINQ, JOB_Sales, MORTDUE, DEBTINC, and DELINQ. And also features negatively affecting Defaulter status: VALUE, LOAN, JOB_ProfExe, JOB_Other, REASON_HomeImp, JOB_Office, CLNO, CLAGE, and YOJ.
- Total Debt-income ratio is the most important feature followed by Credit Delinquency and amount of Derogatory reports.
- Clientele higher DEBINC , higher DELIQ , higher DEROG amounts are an indication that they might default on their mortgage loans.

2. Comparison of various techniques and their relative performance: How do different techniques perform? Which one is performing relatively better? Is there scope to improve the performance further?

- The Logistic Regression model performed similarly in the train (91% precision and accuracy) and test (90% precision and accuracy) datasets.
- LogReg juxtaposed features that positively affect Defaulter status: DEROG, JOB_Self, NINQ, JOB_Sales, MORTDUE, DEBTINC, and DELINQ. And also features negatively affecting Defaulter status: VALUE, LOAN, JOB_ProfExe, JOB_Other, REASON_HomeImp, JOB_Office, CLNO, CLAGE, and YOJ.
- The Decision Tree model is giving a 100% result on the training data, testing data's recall has decreased, testing data's precision has increased. Thus the model is able to identify clients who are at risk of defaulting.
- Decision Tree Tuned gave very same results to previous Decision Tree model.
- Random Forest model did better than both Decision Tree and Decision Tree Tuned.
- Random Forest with class weights model performed slightly better than previous model.
- Random Forest Tuned performed the best so far with testing precision of 1 and accuracy 99%.

3. Proposal for the final solution design: What model do you propose to be adopted? Why is this the best solution to adopt?

- Grid Search aims to choose the best parameters but is dependent on model classifiers on output.
- Tuning the different hyperparameters of each model could improve data's fit.

- Additionally, performing the “boosting” or “bagging” methods may provide greedier approaches to achieve a more accurate model.