Agenda

* More variable selection method
  + Random Trees (joe)
  + Information Value(Josh)
* Modelling (Dom)
  + Find KS value
  + Test for overfitting
* Visualization (vin)
  + Tables
  + Plots
* Finalise?
* What to do include in presentation
  + Introduction
    - Objectives (SMART)
      * What are appropriate approaches to modelling credit risk and what is the current state-of-the-art in this arena? (literature review)
      * Context,
    - Data
      * Which data has been converted
  + Variable Selection
    - VIF
    - Variable Importance
    - Random Forest
    - How we treat the NAs
      * remove (explain)
      * Put into zeros (make them into binary)
    - Correlation between covariates ( removed variables)
    - New Variables
      * Dif\_cred\_len
    - **What are the important variables** in this model and how do they **compare** to variables the bank has found to be traditionally important in its own modelling?
  + Cross Validation
    - Kfold - 10, rep = 3
  + Model Selection
    - Compare a few?
  + Discussion
    - 2. How does this new model perform compared to the one you used previously? How can it be expected to perform on new loans? There are some performance benchmarks available in the project folder on Blackboard.
      * ROC
      * AUC/Gini
      * Accuracy (confusion matrix)
    - 4. What assurances and justifications can you make about the statistical rigor of your model and modelling methodology?
      * Optimum Cut Off
      * Overfitting
  + Conclusion
* Presentation slide
* Question
  + Funded amount ( this variable is known after how much actually the bank is giving )
  + Correlation - 85% ( it depends what u set ur threshold, after removed check the performance of the model )
  + log(ym/(1-ym))
  + Data Splitting before exploration or after (after)
* Todo before Friday
  + Different training set & Cross Validation & different Fit for each model - Joe, Josh
  + Moves Joe’s code to the final code - Joe
  + Exploratory plot - Dom
    - Whatever is not being removed in the beginning.
  + Coefficient Table Single GLM - Vinnie
    - P-value - not significant
    - keep/remove - if coef near zero remove
  + Send Literature review to Dom - Everyone
  + Anova - StepAIC, RandomForest, Lasso - Friday
  + ROC Plot - Friday
  + Accuracy Table - Compare - Friday
  + Break down the presentation