Agenda:

1. Code update from each team member
   1. Vin
      1. Reorganised the code
      2. Update Introduction
      3. Remove loan\_status vs repay plot
      4. Comment out addr\_state plot
      5. Change emp\_leng plot to empl\_length\_factor
2. From the last meeting
   1. Finalised Variable
      1. Variables
         1. Remove the loan amount (funded amount is the amount that the person received)
            1. Can be removed - proved by correlation plot (from Josh)
         2. Remove funded\_amount\_inv , installment, total\_payment, total\_payment\_inv, total\_rec\_prncp,total\_rec\_int
            1. Because it is correlated to loan\_amnt
         3. Remove high correlated variable:
            1. int\_rate and grade
            2. Term and int\_rate
            3. Revol\_util and revol\_bal
         4. Covert employment length into factor
         5. Remove zip\_code
            1. Same purpose as addr\_state
         6. Remove addr\_state
            1. Geographical bias
         7. Remove earliest cr\_line and last credit pull
            1. Make it into credit length
         8. Remove months\_since\_last\_delinq
            1. Same purpose as delinq\_2 years since we changed them into binary
         9. Remove open\_acc
            1. People might open a lot acc for business or sthg
            2. Find literature review to support
         10. Pub\_rec - binary
         11. Remove total\_acc
             1. Not significant from GLM
         12. Remove total payment\_inv
             1. Sub part of total payment
         13. Remove recoveries
             1. Insignificant variables
         14. Remove last\_payment date, last\_payment\_amount
             1. Insignificant variables
         15. Remove last credit pull, earliest credit line
             1. Insignificant variable
         16. Coefficient zero
             1. Issue date
             2. Zip code
             3. Months since last delinq
             4. Last payment date
             5. Last payment amt
             6. Last credit pull date
         17. Insignificant variable
             1. Earliest credit line
             2. Total amt
             3. Recoveries
         18. Remove loan\_status
   2. Finalised Model(Wednesday)
      1. Come out with three and choose?
   3. Validate model
      1. Test,train predict
      2. Cv.glm to do Kfold
   4. Question to ask BoQ
      1. What is loan amount and what is funded amount
      2. What are we comparing our result to?
3. Question from vin:
   1. check repay amount?
      1. Checking how many zeros and ones in data
   2. Line 250 what is defaulted.repay?
      1. Delete