Variables Breakdown

|  |  |
| --- | --- |
| X | What is this |
| ID |  |
| Member\_ID |  |
| Loan amount |  |
| Funded amount |  |
| Funded amount inv |  |
| term |  |
| int\_rate |  |
| installment |  |
| grade |  |
| subgrade |  |
| emplength |  |
| Home ownership |  |
| Annual income |  |
| Verification status |  |
| Issue date |  |
| Loan status |  |
| purpose |  |
| zip\_code |  |
| addr\_state |  |
| dti |  |
| delinq\_2yrs |  |
| earliest \_cr\_line |  |
| Inq last 6 months |  |
| Months since last delinq |  |
| Open acc |  |
| Pub rec |  |
| Revol bal |  |
| Revol utilization |  |
| Total acc |  |
| Total pymnt |  |
| Total payment inv |  |
| Total rec principle |  |
| Total rec interest |  |
| recoveries |  |
| Last pymnt date |  |
| Last payment amount |  |
| Next payment date |  |
| Last credit pull date |  |
| Repay fail |  |

* For the column (Months since delinq), team has to decide whether to put high numbers for NA , or make it into factor (0,1)
* Dti is the ratio of loan/income
* Team has come to an agreement to delete all the N/A rows from the observations (except the next payment date as well as months since delinq)
* Please find the Rmd file (remove\_variables) for the final\_data which exclude all the red variables that have been agreed to removed.

Exploratory Data Question

1. Any interaction
   1. Assume higher income == less likely to default
   2. Does the original state of the debt holder affects the default
   3. Our Assumption a person who has a high loan amt with a low annual income will most likely to default (based on DTI. plot against customer ID and DTI , Group by default output
2. Done by Josh

* Verification Status against repay fail
* Term against repay fail
* Grade / repay fail
* Employment period length vs repay fail