There are files given: train, test and submission. Your submission file must adhere to format specified in the given submission file. This data set comprises of information captured in December 2016. Following is the description of variables given:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Description** | **Column Type** | **Insights** |
| member\_id | unique ID assigned to each member |  | Chuck coumn |
| loan\_amnt | loan amount ($) applied by the member | Number |  |
| funded\_amnt | loan amount ($) sanctioned by the bank | Number |  |
| funded\_amnt\_inv | loan amount ($) sanctioned by the investors | Number |  |
| term | term of loan (in months) | Number |  |
| batch\_enrolled | batch numbers allotted to members |  | Chuck  -see if any insights can be derived |
| int\_rate | interest rate (%) on loan | Float |  |
| grade | grade assigned by the bank | Factor |  |
| sub\_grade | grade assigned by the bank | Factor | Correlation |
| emp\_title | job / Employer title of member |  | Chuck  -Include later in model |
| emp\_length | employment length, where 0 means less than one year and 10 means ten or more years | Number |  |
| home\_ownership | status of home ownership | Factor |  |
| annual\_inc | annual income ($) reported by the member | Number |  |
| verification\_status | status of income verified by the bank | Factor |  |
| pymnt\_plan | indicates if any payment plan has started against loan | Factor |  |
| desc | loan description provided by member |  | Chuck  But classify |
| purpose | purpose of loan |  | Chuck |
| title | loan title provided by member |  |  |
| zip\_code | first three digits of area zipcode of member | Number |  |
| addr\_state | living state of member |  | Chuck  As Zipcode is almost same as Addr\_State |
| dti | ratio of member's total monthly debt repayment excluding mortgage divided by self reported monthly income | Numeric |  |
| delinq\_2yrs | number of 30+ days delinquency in past 2 years | Number |  |
| inq\_last\_6mths | number of inquiries in last 6 months | Number |  |
| mths\_since\_last\_delinq | number of months since last delinq | Number |  |
| mths\_since\_last\_record | number of months since last public record | Number |  |
| open\_acc | number of open credit line in member's credit line | Number |  |
| pub\_rec | number of derogatory public records | Number |  |
| revol\_bal | total credit revolving balance | Number |  |
| revol\_util | amount of credit a member is using relative to revol\_bal | Number |  |
| total\_acc | total number of credit lines available in members credit line | Number |  |
| initial\_list\_status | unique listing status of the loan - W(Waiting), F(Forwarded) | Factor |  |
| total\_rec\_int | interest received till date | Number |  |
| total\_rec\_late\_fee | Late fee received till date | Number |  |
| recoveries | post charge off gross recovery | Number |  |
| collection\_recovery\_fee | post charge off collection fee | Number |  |
| collections\_12\_mths\_ex\_med | number of collections in last 12 months excluding medical collections | Number |  |
| mths\_since\_last\_major\_derog | months since most recent 90 day or worse rating | Number |  |
| application\_type | indicates when the member is an individual or joint | Factor |  |
| verification\_status\_joint | indicates if the joint members income was verified by the bank | Factor |  |
| last\_week\_pay | indicates how long (in weeks) a member has paid EMI after batch enrolled | Number |  |
| acc\_now\_delinq | number of accounts on which the member is delinquent | Number |  |
| tot\_coll\_amt | total collection amount ever owed | Number |  |
| tot\_cur\_bal | total current balance of all accounts | Number |  |
| total\_rev\_hi\_lim | total revolving credit limit | Number |  |
| loan\_status | status of loan amount, 1 = Defaulter, 0 = Non Defaulters | Factor | Target Variable |

**Evaluation Metric**

Submissions will be evaluated based on AUC-ROC score. For more information about this metric, [read here](https://en.wikipedia.org/wiki/Receiver_operating_characteristic).

**Update (17th March, 3pm)** - If you've made any earlier submission based on logloss, kindly optimize it for AUC-ROC score. All the existing submissions will be re-evaluated using the AUC-ROC scoring criteria.