Data Collection and Description

The dataset collected is Primary and Secondary data from chit-fund companies for the Credit Scoring project in India (2012-12-01), The data set includes 5 tables, but because of the analysis problem, I only took 3 tables including cf1_delhi_collateral, cf1_delhi_surety, cf1_delhi_transaction_data. Table cf1_delhi_collateral contains information indicating members' collateral assets, each member can have more than 1 collateral asset. Table cf1_delhi_surety shows detailed information about each member's guarantor, including information such as age, gender, occupation, land assets, house assets, etc. Table cf1_delhi_transaction_data contains information about transactions of each member in the chit fund, including information about transaction date, chit value, payment method, payment nature (early, late, partial), and participation auction. The following table is a detailed description of each column in each table

Variable	**Data type**	**Description**
Table: cf1_delhi_collateral		
chit_id	categorical	Identifier for individual members of each chit fund.
p_recno	categorical	The receipt number of the payment a member contributed to the chit fund.
collateral	categorical	The type of collateral(s) the members has (e.g., CER, CHT, etc.).
chit_value	real-valued multiplicative	The total value of a chit fund is obtained by (duration x monthly_contribution)
duration	count	The duration of operation for a chit fund, must be equal to the number of members involved.

monthly_contribution	real-valued multiplicative	The amount of money each member must contribute to the chit fund every month
year	categorical	The inception year of a chit fund.
Table: cf1_delhi_surety		
chit	categorical	Identifier for each chit fund.
chit_id	*Referenced from table cf1_delhi_collateral*	
p_recno	*Referenced from table cf1_delhi_collateral*	
winning_aucn	real-valued multiplicative	The highest discounted price that the member to win the bid
duration	*Referenced from table cf1_delhi_collateral*	
fman	binary	If a foreman manages the chit fund, with '0' for no and '1' for yes.
surety_p_recno	categorical	Identifier for a guarantor associated with each chit fund member's contributions.
n_surety	count	The number of guarantors associated with each chit fund member's contributions.
age	count	Guarantor’s age.
salary	real-valued multiplicative	The amount of money the guarantors receives as salary on a monthly basis.
sex	binary	Guarantor’s gender (“M”: “male”, “F”: “female”).
occupation	categorical	Guarantor’s occupation or job title (“B”: “Business”; “GS”: “Graduate Student”; “HW”: “Housework”; “P”: “Professor”; “PS”: “Public Service”; “R”: “Researcher”; “SE”: “Software Engineer”.

other_chits	binary	If a guarantor associated with two or more chit in a same year (“Yes”: “1”; “No”: “0”).
surety_others	binary	If a guarantor associated with two or more members in a same year (“Yes”: “1”; “No”: “0”).
years_of_service	real-valued multiplicative	The number of years a Guarantor has been working or serving in a occupation.
house_owner	binary	If a guarantor owns the house (“Yes”: “1”; “No”: “0”).
land_owner	binary	If a guarantor owns the land (“Yes”: “1”; “No”: “0”).
income_tax	binary	If a guarantor is required to pay income tax or not. (“Yes”: “1”; “No”: “0”).
insurance_policy	binary	If a guarantor has an insurance policy or not (“Yes”: “1”; “No”: “0”).
insurance_amount	real-valued multiplicative	The amount of money that guarantor will receive in case of an insurance event such as accidents, illnesses, death, etc.
res_pin	categorical	Reserve Participant Identification Number of a members
off_pin	categorical	Official Participant Identification Number of a members
chit_value	*Referenced from table cf1_delhi_collateral*	
monthly_contribution	*Referenced from table cf1_delhi_collateral*	
year	*Referenced from table cf1_delhi_collateral*	

Table: cf1_delhi_transaction_data

chit	*Referenced from table cf1_delhi_surety*
chit_id	*Referenced from table cf1_delhi_collateral*

p_recno	*Referenced from table cf1_delhi_collateral*	
winning_aucn	*Referenced from table cf1_delhi_surety*	
aucn_no	categorical	Auction Number is an identifier assigned to each monthly auction event within the chit fund.
aucn_date	categorical	Auction Date would represent the date the auction takes place within the chit fund.
inst_due	real-valued multiplicative	The installment that is currently due which members making regular contributions monthly.
inst_paid	real-valued multiplicative	The installment amount that the member has paid during the month
inst_spread	real-valued multiplicative	The spread between the installment due and the outstanding installment (The outstanding installment is calculated by: total_inst_due - total_inst_paid).
total_inst_due	real-valued multiplicative	The cumulative amounts of installments that are currently due from each of the members in the chit fund.
total_inst_paid	real-valued multiplicative	The cumulative amounts of installments that the member has paid in the chit fund.
div_due	real-valued multiplicative	The dividend that is currently due is received by members in each auction (The dividend is calculated by: monthly_contribution - inst_due).
div_paid	real-valued multiplicative	The dividend is based on what the member has paid during the month

total_div_due	real-valued multiplicative	The cumulative amounts of dividend due that are currently due from each of the members in the chit fund.
total_div_paid	real-valued multiplicative	The cumulative amounts of dividend paid that are currently due from each of the members in the chit fund.
participation	binary	If the member participated in the auction during the month (“Participate”: “1”; “Not Participate”: “0”).
all_bids	real-valued multiplicative	All Bids (Bid is discount price offered) would encompass the individual bids submitted by members participants during an auction.
win_loss	binary	If the member participants win the bid (“Yes”: “1”; “No”: “0”).
win_bid_amt	real-valued multiplicative	The amount of discount price the member bid to win the Bids
prized_amt	real-valued multiplicative	The amount of prize to the members who win the Bids (Prized Amount is calculated by: chit_value - prized_amt).
chit_value	*Referenced from table cf1_delhi_collateral*	
start_date	categorical	The date of starting paying a monthly installment
monthly_contribution	*Referenced from table cf1_delhi_collateral*	
duration	*Referenced from table cf1_delhi_collateral*	
month	categorical	Month of operation of chit fund
tot_memb	count	The total number of members of a chit-fund

fman_tkt	binomial	The management fee that the foreman receives in tickets during the month is approved by the total number of members in the fund
bylaw_no	categorical	The Law is applied by Chit fund
penalty	real-valued multiplicative	Penalty refers to a financial charge imposed on members for non-compliance with the agreed-upon terms and conditions.
postage_cost	real-valued multiplicative	The cost of postage may include the charges for sending communication, documents, or notices to the members via postal services.
nj_stamp_cost	real-valued multiplicative	The cost of stamp.
other_cost	real-valued multiplicative	The cost other may include: related to legal compliance, documentation, or regulatory requirements, cost specific software for its operations.
by_chq	binary	If members make installment payments by cheque (“Yes”: “1”; “No”: “0”).
by_cash	binary	If members make installment payments by cash (“Yes”: “1”; “No”: “0”).
by_other	binary	If members make installment payments by others (“Yes”: “1”; “No”: “0”).
bounced_chq	binary	If the member is bounced the cheque (“Yes”: “1”; “No”: “0”).
last_trans_date	categorical	The date of the last transaction made by the member in the month

last_payment_date	categorical	The date of last paying a monthly installment
missed_inst	binary	If the member does not make any transactions during the month (“Yes”: “1”; “No”: “0”).
missed_div	binary	If the member does not receive profit during the month (because no members bids) (“Yes”: “1”; “No”: “0”).
diff_inst	real-valued additive	Installment difference between total installment due and total installment paid (Installment difference is calculated by: total_inst_due - total_inst_paid).
no_trans	count	Number of transactions made by members during the month.
total_trans	count	The cumulative amounts of transactions
multi_payment	binary	If the member makes payments using multiple methods (“Yes”: “1”; “No”: “0”).
early_payment	binary	If members make payment transactions before the last date (“Yes”: “1”; “No”: “0”).
part_payment	binary	If members make payment transactions or not which the cumulative amounts of installments due are larger than the cumulative amounts of installments paid (total_inst_due > total_inst_paid) is considered a partial payment (“Yes”: “1”; “No”: “0”).
irr_payment	binary	If the members make payment transactions irregular (“Yes”: “1”; “No”: “0”).

late_payment	binary	If members make payment transactions after the last date (“Yes”: “1”; “No”: “0”).
default	binary	If a member default to fulfill their financial obligations as per the terms and conditions outlined in the chit agreement during the month (“Yes”: “1”; “No”: “0”).
monthly_income	real-valued multiplicative	The amount of income a member receives monthly
sex	binary	Member’s gender (“M”: “male”, “F”: “female”).
age	count	Member’s age.
occupation	categorical	Member’s occupation
lottery	binary	If an auction must be decided by lottery to see which member wins the Bids (“Yes”: “1”; “No”: “0”).
bid_type	categorical	Determine the bid type by counting how many people are participating in the auction. • “0”: “There are no members participating in the auction” • “1”: “There is one member participating in the auction” • “2”: “There is more than one member participating in the auction”.
before_after	binary	• The transaction is made **before** the auction is successful (win_loss = 1): “0”.

		• The transaction is made **after** the auction is successful (win_loss = 1): "1".
all_trans	categorical	Count all transactions
default_90	binary	If the member defaults continuously for 3 months ("Yes": "1"; "No": "0").

Table 3. 1: Describe all the variables of the 3 collected tables