GLOBAL ACADEMIC RESEARCH INSTITUTE

COLOMBO, SRI LANKA

International Journal of Management (IJM)

ISSN 2424-6433

Volume: 01 | Issue: 01

On 31st March 2016

http://www.research.lk

Author: Anjali Sharma, Ambudheesh Parasar

Chandragupt Institute of Management Patna, India

GARI Publisher | Finance Management | Volume: 01 | Issue: 01

Article ID: IN/GARI/ICBME/2016/202 | Pages: 22-53 (31)

ISSN 2424-6433 | Edit: GARI Editorial Team

Received: 03.02.2016 | Publish: 31.03.2016

"An Analysis of Mortgage Loan Approval with Customer Segmentation"

Anjali Sharma (anjali14@cimp.ac.in)

Chandragupt Institute of Management Patna, India

Ambudheesh Parasar (ambudheesh@gmail.com)

Independent Researcher

ABSTRACT

A Mortgage loan is a secured loan on the borrower's property. Of late, there has been numerous case of default in the Mortgage loan which significantly increases the Non-performing asset of the financial institutions like banks. This study examine the importance of factors such as age, number of dependents, time lag, tenure, Credit Information Bureau India limited (CIBIL) score, interest rate, loan, per capita residual income, education, gender and employment type on granting the Mortgage Loan. The purpose of this study is to find out the critical factors from the above mentioned factors to develop a mathematical model which can help in identifying and differentiating a possible defaulter from non-defaulter. This study also analyses the segmentation of the customer based on the factors. The data for this study collected from one of the leading financial institute of India. Stratified random sampling is used for the sample selection process. The study attempts to develop a mathematical model for granting the loan which help in reducing the non-performing asset of the financial Institution.

Key Words: Mortgage Loan, Factors, Discriminant analysis, Segmentation.

INTRODUCTION

Mortgage is based on the fundamental concept of basic amenities of the life which is having a shelter i.e. a house. Mortgage is destined to permit individuals to fund their ownership for their houses. A Mortgage is a credit offered out to people by a bank or other lending institutions intended for the purchase of a house. It is a security supported loan, implying that when an individual goes into a bank to get a mortgage, the bank will own the house, and they will utilize that house as collateral for their credit.

Area of the study is Kolkata, which is one of the metro cities in India. The study explores the problem of mortgage loan default. Loan Default means that the debt holder has not met his obligation for the payment of the Equated Monthly Installment (EMI).Three consecutive defaults in the payment of EMI's leads to Non-Performing Assets (NPA). The very concept of Non Performing Asset (NPA) is limited to loans, advances and investments. Till the time the asset generates the income anticipated, it is treated as a performing asset, but when it's unable to do so then it becomes a "Non-Performing Asset", which is a great concern for the financial institutions. According to Reserve Bank of India, the performance of the financial sectors has further worsened in the last one year, with gross non-performing resource (GNPA) proportion creeping to 4.45 per cent as on March 15 Last year, as compared to 4.1 per cent in March 2014.

This study mainly comes up with a model for the financial institution which can help them to anticipate their future Non-Performing Assets and help them to reduce the defaulters with the support of factors such as Age of the Borrower, Number of dependents on the borrower, his/her time lag, tenure, Interest Rate, Per capita residual income, CIBIL score, loan taken by the borrower, Gender, Education qualification and his/her source of income.

Objectives of the study:

The objective of this study is to assess the following points:

a) Impact of borrower's Profile on loan default.
b) Identification of critical factors for Assessing Mortgage Loan.
c) Predictive Model for judging Defaulters and Non-defaulters.
d) Segmenting the Customer base.

LITERATURE REVIEW

Mortgage

Mortgages are amongst the most fundamental monetary instruments on the grounds that they are so prevalent. Mortgage moneylenders have sprung up in the world over in light of the fact that in today's market, individuals are essentially not able to purchase a home without having a mortgage to do as such. This implies that individuals are kept from continually owning a home on the off chance that they don't meet all requirements for a mortgage. This additionally implies that individuals won't have the capacity to satisfy their objectives of owning a home on the off chance that it was not for a mortgage. A mortgage is an exceptionally liberating idea in light of the fact that it permits individuals to get the home that they need and to achieve their objectives through the assistance of a bank and the mortgage that they provide.

Default

Default happens when a debt holder has not met his or her legitimate commitments as per the obligation contract. For instance an indebted person has not paid the installment, or has disregarded an advance pledge (condition) of the obligation contract (Ameyaw-Amankwah, 2011). A default is the inability to pay back a loan. Default may happen if the debt holder is either unwilling or not able to pay their obligation. A loan default happens when the borrower does not pay obliged installments or in some other way does not consent to the terms of a credit. (Murray, 2011).In addition, Pearson and Greeff (2006) defined default as a danger edge that portrays the borrower's repayment history where he or she missed no less than three instalments in a 24 month period. This speaks to a point in time and indicator of conduct, wherein there is an obvious increment in the risk that the borrower in the long run would default, by stopping all repayments. The definition is reliable with international standards, and is essential in light of the fact that steady investigation obliged a typical definition. This definition does not imply that the borrower has totally stopped paying the loan and consequently been allude to collection or legal procedures; or from an accounting viewpoint that the loan had been characterized as bad or doubtful, or written-off. Credit default can be characterized as the failure of a borrower to satisfy his or her loan commitment as and when due (Balogun & Alimi, 1990).

Causes for Default

The reasons for default might be many folds. As Ahmad, (1997) argued that reasons for loan default incorporate; absence of readiness to pay credits combined with redirection of funds by borrowers, willful negligence and ill-advised evaluation by credit officers. (Balogun E. a., 1998) additionally recognized the significant reasons for credit default as advance deficiencies, delay in time of advance delivery, small business holding, high interest rate, age of client, poor supervision, non profitability of farm businesses and undue government intervention with government sponsored credit programs. Also, Akinwumi and Ajayi (1990) figured out that farm size, family, scale of operation, family livelihood expenses and introduction to sound management strategies are some of the components that can impact the repayment limit of client. As indicated by Olomola (1999), loan disbursement time lag and high interest rate can altogether expand transaction cost and can likewise antagonistically influence repayment execution. After an intensive survey of the various banks in India, Berger and De Young (1995) distinguished the primary cause of default of credits from industrial sector as wrong choice of the entrepreneur, lacking analysis of venture reasonability, insufficiency of security/equitable mortgage against loans, farfetched terms and schedule of repayment, absence of follow-up measures and default because of normal disasters. The study directed by Okorie (1986) in Ondo state in Nigeria uncovered that the nature, time of disbursement, supervision and productivity of organization, added to the repayment capacity and subsequently high default rates. Other discriminating components connected with loan default are: kind of the credit; term of the credit; interest rate on the credit; poor credit record; borrowers' salary and the transaction cost.

Studies identified household and individual characteristics for loan default summarized below

Age and Loan default

Different studies identified ‘age’ as one of the important factor in loan default. Capozza et al. (1997) (Capozza, Kazarian, & Thomson, 1997; Itoo , Mutharasu, & Filipe, 2013) demonstrated that the borrower's age is adversely connected with the default possibility. Hakim and Haddad (1999) study shows that the age of the borrower is essentially adversely correlated with the default likelihood. (Jacobson & Roszbach, 2003) found that the applicant's age is fundamentally contrarily corresponded with the loan default. Cairney and Boyle (2004) demonstrate that the age of the borrower is negatively associated with the loan default. However, Kumar (2010) found there is no relation between the age of the borrower and mortgage defaults.

Loan amount and Loan default

Paul Bennett et al. (1997) studied that loan size is adversely related with the mortgage defaults. Hakim and Haddad (1999) considered the impacts of the borrower's traits and the credit qualities on the mortgage loan default with the help of failure-time model. Their outcome demonstrates that the loan size is adversely related with the default likelihood.

Gender of borrower and Loan default

Jacobson and Roszbach (2003) showed that the applicant's gender adversely corresponded with the loan default.

Borrower's educational qualification and Loan default

Liu and Lee (1997) examines that the borrower's educational qualification is negatively related with the mortgage loan default, which was substantiated by Cairney and Boyle (2004), who indicated the education level has an adverse relation with the loan default risk.

CIBIL Score and Loan Default

CIBIL remains for Credit Information Bureau of India, an association which gives a full picture of one's reimbursement track and the credit responsibility. The Credit Information Bureau of India (CIBIL) gives score between300 to 900. The lower the credit score, the more chances of loan rejection.

Galindo A. Miller M.J., (2003) said that "the accumulation and support of satisfactory positive information will essentially build specialized, and monetary necessities of credit agencies, raising the expense for credit organizations, which will at last be reflected in the expense of credits for customers", presuming that "shopper data, dependable loaning practices and the legitimate environment should be adjusted in any open approach technique".

Interest Rate and Loan default

(Campbell & Dietrich, 1983) showed that the interest rates significantly explain mortgage prepayment, delinquencies and defaults. (Har & Eng, 2004) found the loan interest rate was significantly positively correlated with the mortgage loan default. However, (Teo & Ong)(2005) demonstrated that the interest rate is adversely connected with the home loan credit default

RESEARCH METHODOLOGY

This section consists of Objectives of the study, Data Collection, Sample size determination, Statistical tools, Demographic profile of the respondent and the Methodology.

Data Collection

The study is based on data extracted from one of the leading financial institutes of India from secondary sources maintained in the bank but not published. In order to understand the reliability and Suitability of the data one of the researchers has spent 60 days and the researcher recorded the data from the borrower's file. The Borrower's are selected in such a way that the number of defaulters is more or less equal to the number of non-defaulter.

Sample Size Determination

Target Population-The Customers who have applied for the loan from the financial Institution.

For sample size determination, following values are used

Level of Confidence at 99%, Margin of error 4%, Probability of paying back loan as 0.98 and probability of not paying back loan as 0.02

So, mathematically sample size is determined to be

S= (2.57/0.04)^2*0.98*0.02

= 80.9

So the sample size is taken as 81.

Sampling Methodology:

The sampling method is Stratified Random sampling, since the data are geographically separated and demographically arranged. The Strata is age and education qualification of the Borrower's.

Statistical Design:

The researchers use the following tools for the analysis of Data Viz., Chi square, Factor Analysis, Discriminant Analysis and cluster analysis. SPSS 20 is used for the analysis of the above statistical tool and tabulation of Processed data.

Methodology

For objective 1 which is to assess the impact of borrower's profile on the loan default, chi-square test of association is used. Chi square test of association is being used to find out the strength of relationship between Borrower's profile which included factors like number of dependents, gender, source of income and educational qualification.

For objective 2 which is to assess the critical factors, factor analysis is performed over 8 variables namely age, time lag, tenure, PCRI, CIBIL score, interest rate, loan and educational qualification. Before the application of factor analysis, normalization is done using z score. Factor analysis helps in identifying the critical factors and that becomes the base for further research.

For objective 3 which is to assess a predictive model for judging defaulters and Non-defaulters, Discriminant analysis is performed. Discriminant analysis gave a mathematical model and that model is used to predict whether a borrower would be a defaulter or Non-defaulter.

For objective 4 which is to segment the customer base, Cluster analysis is performed. Cluster analysis helps in segmenting the customer base and on that segments are identified

Demographic Profile of the Respondent

Age

Age is a factor in providing the loan .The minimum age taken for the research work is 20 years whereas maximum age taken for the project is 56.For the purpose of research, 81 applicants age are studied along with the other 10 factors

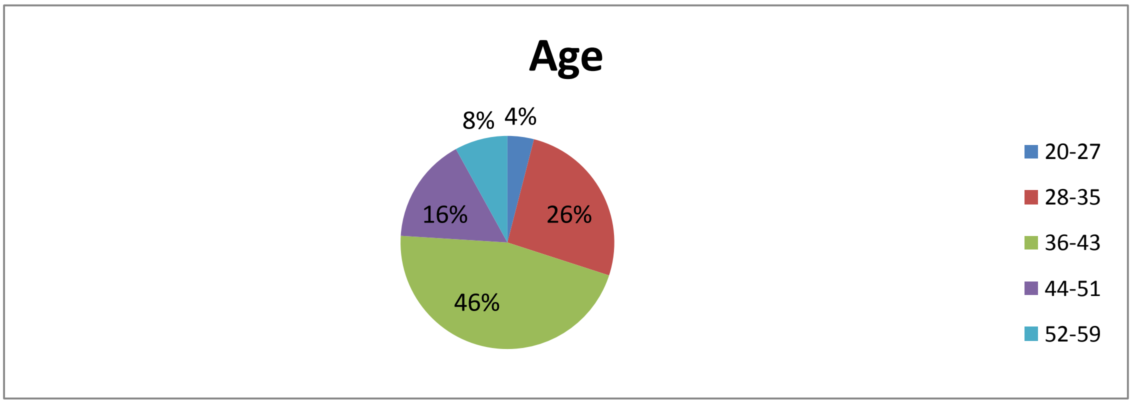

Chart1: Percentage wise distribution of Age

Dependent

Numbers of dependent are a factor in giving the loan. Dependent includes those who are directly dependent on the loan taker. For the purpose of research, Dependent of 81 applicants' is studied along with the other 10 factors. The minimum numbers of dependent are 0 whereas the maximum numbers of dependent are 5.

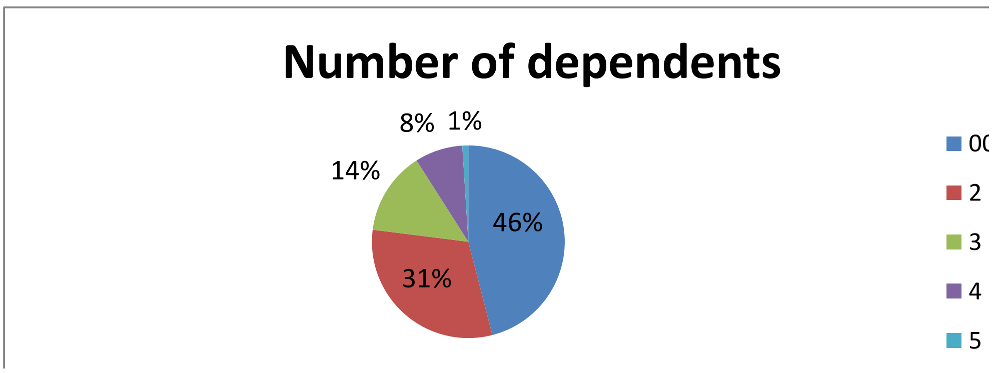

Chart2: Percentage wise distribution of Number of dependents

Time lag

This factor shows the difference in the date of agreement and the date on which the loan disbursement took place. This factor has been defined in number of days. The minimum time lag is 10 days whereas the maximum Time lag is 59.For the purpose of research, 81 applicant's Time lag is studied along with the other 10 factors.

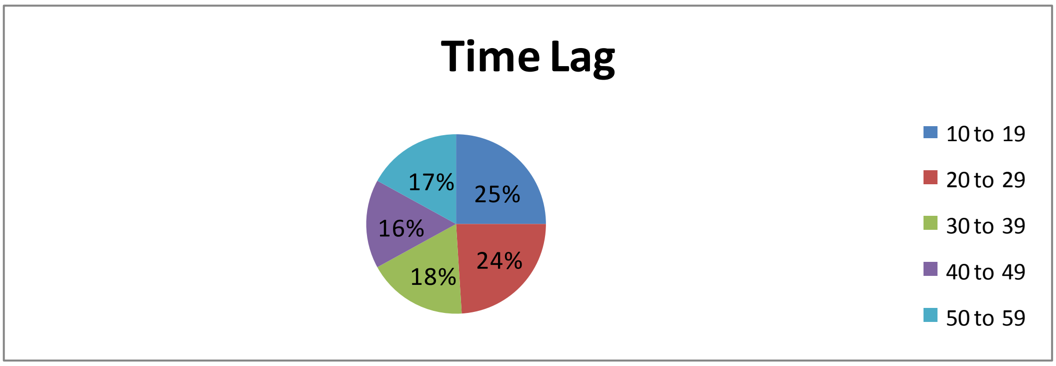

Chart 3: Percentage wise distribution of Time lag

Tenure

This factor shows for how many months the loan is availed and its repayment will take place. Tenure is recorded in months. The tenure period ranges from 48 months to 220 months.

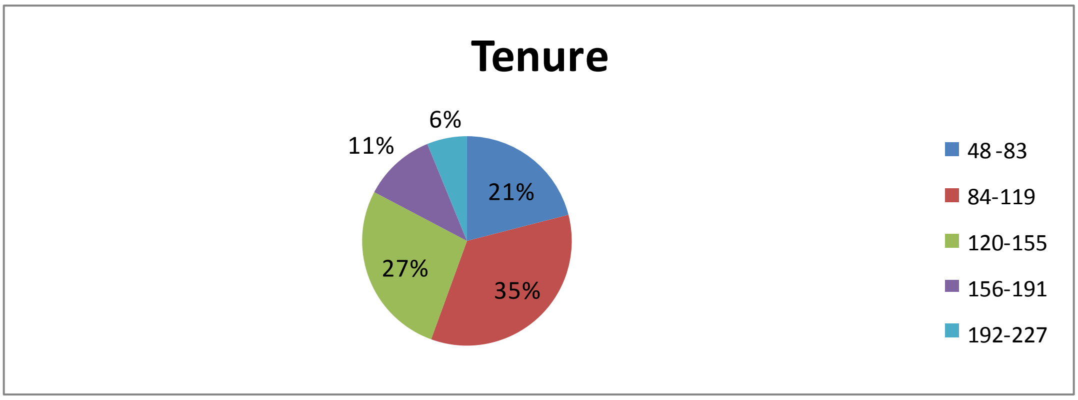

Chart 4: Percent wise distribution of Tenure

Interest Rate

It is the rate at which the repayment of the sanctioned loan would take place. The Interest Rate in the research varies from 13.25 to 14.75.

For the purpose of research, 81 applicant's Interest rate is studied along with the other factors.

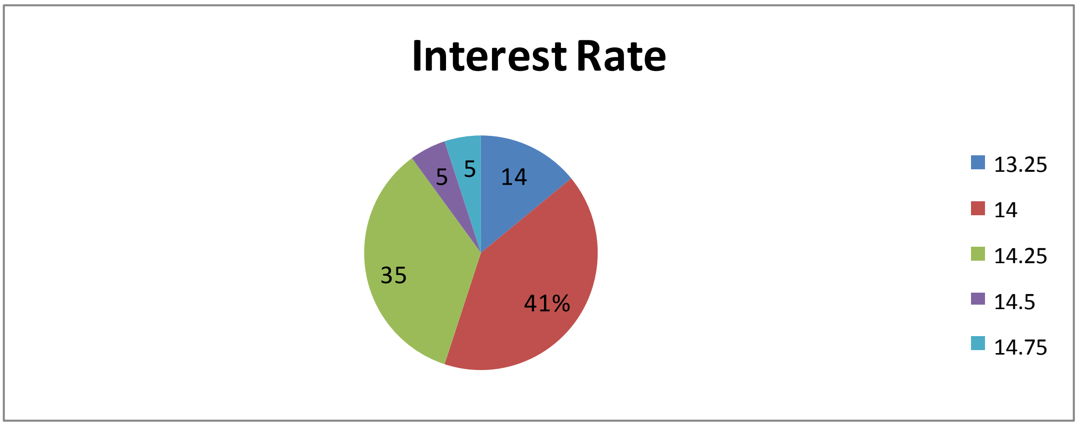

Chart 5: Percent wise Distribution of Interest Rate

Per Capita Residual Income

Residual Income means the amount of income that an individual has after all personal debts, including the mortgage, have been paid. This calculation was usually made on a monthly basis, after the monthly bills and debts are paid. Also, when a mortgage has been paid off in its entirety, the income that individual had been putting toward the mortgage becomes residual income. As the name suggest, PCRI is nothing but the Residual income divided by the no. of Family members. Residual income is often an important component of securing a loan. The loaning institution usually assesses the amount of residual income an individual has left after paying off other debts each month. The Minimum PCRI in the sample of 81 respondents is Rs.1827 whereas the maximum is Rs.22253.

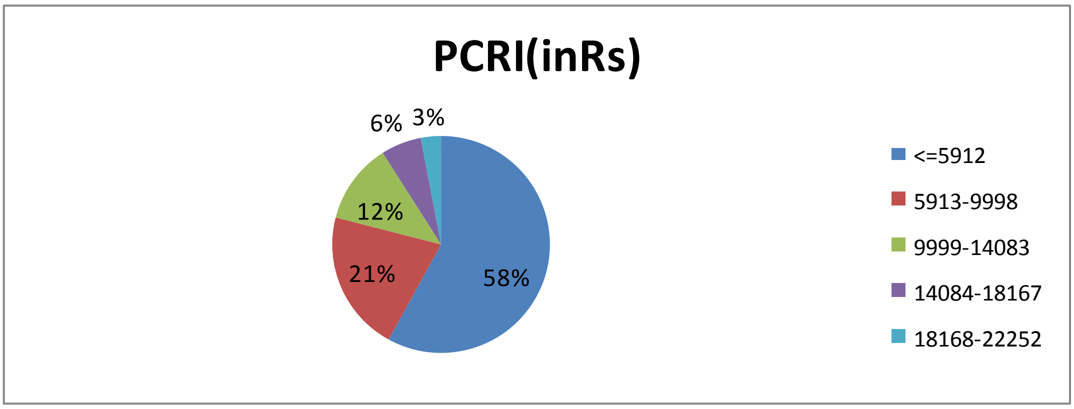

Chart 6: Percentage wise Distribution of PCRI

Loan

The amount of money borrowed from GIC is termed as Loan. The minimum loan taken for the sample is Rs. 3500000 and the maximum amount is Rs. 4450000.

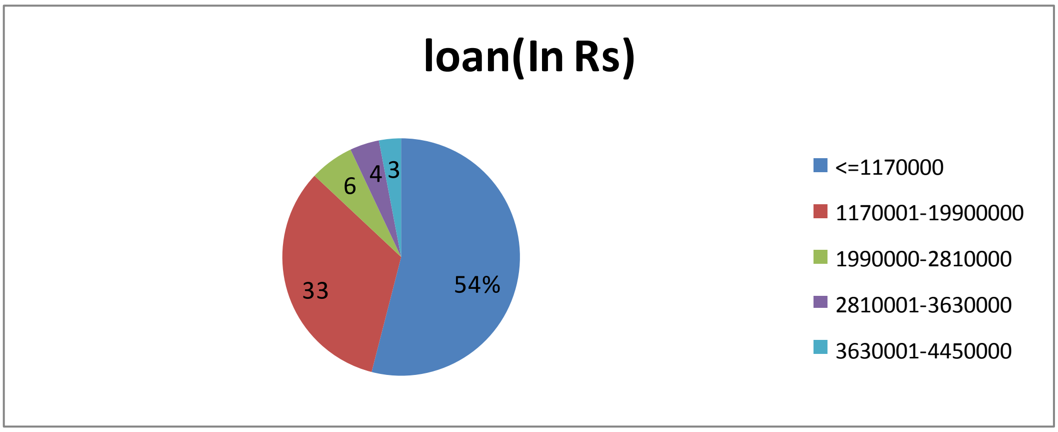

Chart 7: Percentage wise distribution of Loan

CIBIL SCORE

Credit Information Bureau (India) Ltd; CIBIL is India's first Credit Information Company, also commonly referred as a Credit Bureau. CIBIL collect and maintain records of individuals' and non-individuals' (commercial entities) payments pertaining to loans and credit cards. These records are submitted to the CIBIL by banks and other lenders on a monthly basis; using this information a Credit Information Report (CIR) and Credit Score is developed, enabling lenders to evaluate and approve loan applications. A Credit Bureau is licensed by the RBI and governed by the Credit Information Companies (Regulation) Act of 2005.In the sample size of 81 respondents, the lowest CIBIL score is 528 whereas the highest CIBIL score is 872.

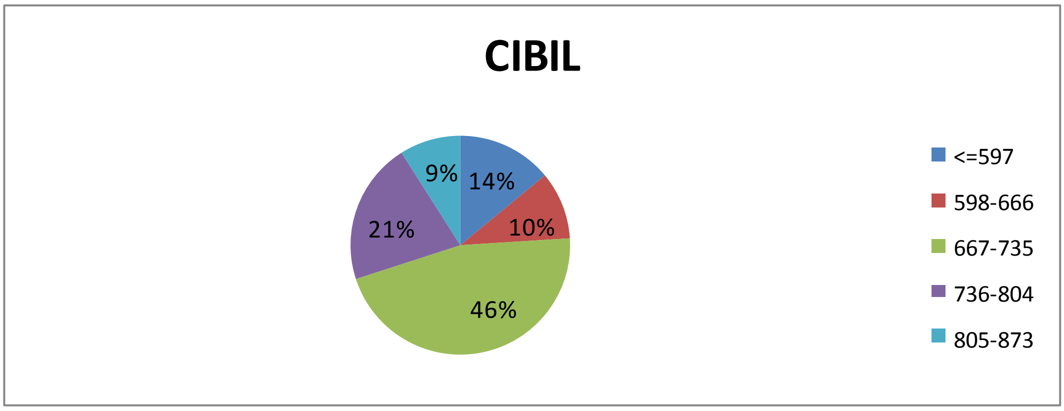

Chart 8: Percentage wise distribution of CIBIL score

Education Qualification

Education qualification is considered to be a factor in Loan default Process. In education qualification 1 denotes the loan getters who has education up to 10^{th} grade,2 denotes those who has an education qualification till intermediate level ,3 denotes those who has an education qualification till graduation where 4 denotes post graduation and 5 denotes doctoral level.

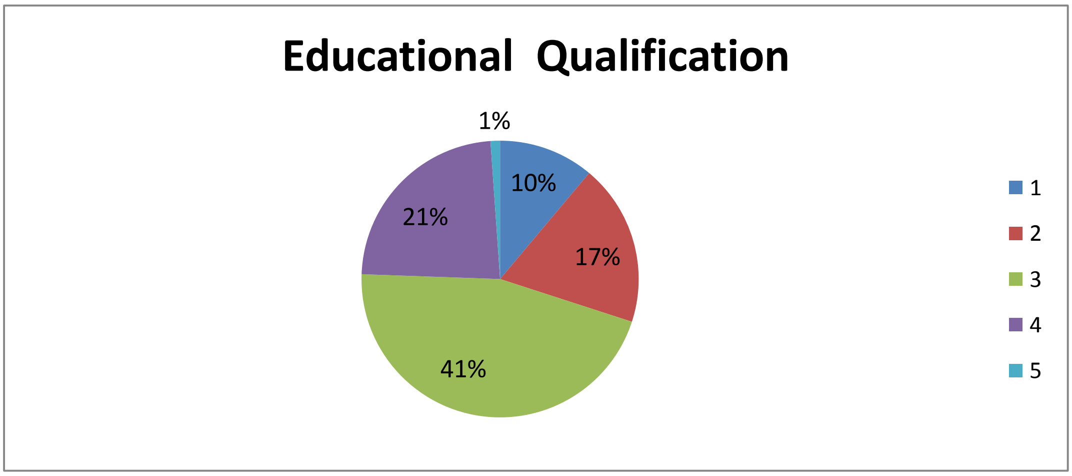

Chart 9: Percentage wise distribution of educational qualification

Gender

1 refers to male in the sample whereas 2 refer to female in the sample. Number of males are 50 whereas number of females are 31.

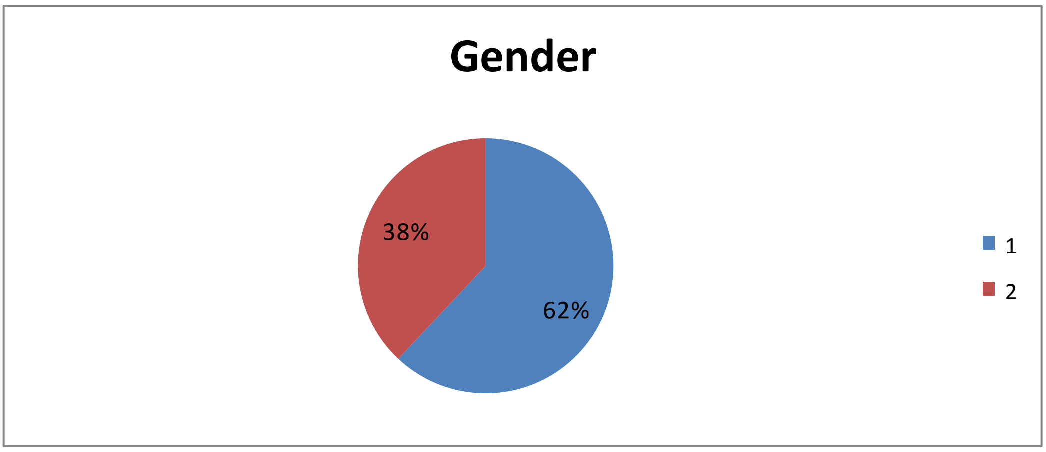

Chart 10: Percentage wise distribution of Gender

Source of Income

1 refers to Businessman/Businesswomen in the sample whereas 2 refer to salaried people the sample. Number of Businessman/Businesswoman is 40 whereas number of salaried people is 41.

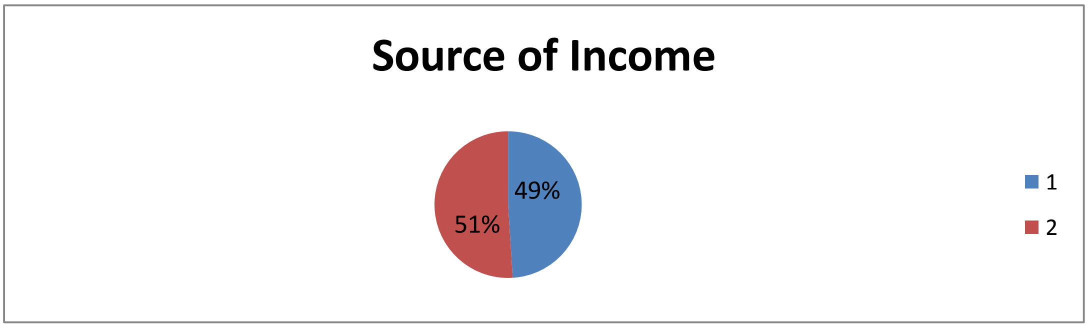

Chart 11: Percentage wise distribution of source of income

DATA ANALYSIS AND DISCUSSION

For the research work following factors are identified for analysis

Based on the literature review and field experience the following factors have been identified which expected to have influence on repayment of loan taken under mortgage scheme.

1. Age
2. Number of Dependents
3. Time lag in Disbursement
4. Tenure of Loan
5. Interest Rate
6. Per Capita Residual Income
7. CIBIL Score
8. Loan Amount
9. Gender
10. Education Qualification
11. Source of Income (Salaried vs. Business)

Association between Borrower's profile and Loan default

The borrower's Profile Information like gender (M/F where male was ranked as 1 and female was ranked as 2), education qualification (where 1- up to 10^{th} grade,2-upto intermediate level,3-upto graduation level,4- Post graduation 5-Phd),number of dependents and source of income (where 1- salaried class and 2-business class) are tested for association with the defaulters list (where 1- defaulter and 2-Non-defaulter)

Association between gender and Loan defaulter's list.

A chi square test of association is performed and no relationship is found between the defaulter's list and gender at the significance level of 0.01. $\chi2$ (1,N=81)=1.045,p=0.307.

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**1.045[a]**	1	.307

Table1: Chi-square test of association between Loan defaulter's list and gender.

Association between Education qualification and Loan defaulter's list.

A chi square test of association is performed and a significant relationship is found between the defaulter's list and Education qualification at the significance level of 0.01. $\chi2$ (4,N=81)=14.985,p=0.005.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**14.985**[a]	4	.005

Table 2: Chi-square test of association between Loan defaulter's list and Education qualification

So, the borrower with higher education qualification tends to default less

Association between Source of income and Loan defaulter's list.

A chi square test of association is performed and no significant relationship is found between the defaulter's list and Source of income at the significance level of 0.01. $\chi 2$ (1,N=81)=0.121,p=0.728.

	Value	df	Asymp. Sig. (2-sided)
Pearson chi-Square	**.121**[a]	1	.728

Table 3: Chi-square test of association between Source of income and Defaulter's list.

Association between Number of dependents and Loan defaulter's list.

A chi square test of association is performed and no significant relationship is found between the defaulter's list and number of dependents at the significance level of 0.01. $\chi 2$ (4,N=81)=5.777,p=0.216.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**5.777**[a]	4	.216

Table 4: Chi-square test of association between number of dependents and Defaulter's list

Identification of critical factors

For the identification of critical factors, following factors are taken into consideration:

- Age of the respondents
- Time taken for granting the loan
- Tenure for the loan
- Interest rate

- PCRI
- Loan
- CIBIL
- Education level

Factor analysis is performed with these 8 factors.

Since the 8 factors are on different scale, so data is normalized using the z score.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.455
Bartlett's Test of Sphericity	Approx. Chi-Square	127.484
	df	28
	Sig.	.000

Communalities

	Initial	Extraction
Zscore (Age)	1.000	.730
Zscore (Time)	1.000	.854
Zscore (Tenure)	1.000	.868
Zscore (Interestrate)	1.000	.560
Zscore (PCRI)	1.000	.784
Zscore (Loan)	1.000	.814
Zscore (CIBIL)	1.000	.768
Zscore (Education)	1.000	.645

Extraction Method: Principal Component Analysis

Table 5: KMO and Bartlett's Test

Prior to the extraction of the factors, several tests are used to assess the suitability of the respondent data for factor analysis. These tests include Kaiser-Meyer-Olkin (KMO), Measure of Sampling Adequacy and Bartlett's Test of Sphericity. The KMO value in this research comes out to be 0.455 which is close to 0.5 so the factors are considered just well enough for the factor analysis.

Bartlett's Test of Sphericity:

The null hypothesis is that the inter-correlation matrix comes from a population in which the variables are non-collinear (i.e. an identity matrix) and that the non-zero correlations in the sample matrix are due to sampling error.

Test Results for Bartlett's test of Sphericity

$\chi 2 = 127.484$

df = 28

$p < 0.001$

Statistical Decision

The sample inter correlation matrix do not come from a population in which the inter correlation

Matrix is an identity matrix.

After these two test,,critical factors are obtained.

Total Variance Explained

Component	Initial Eigenvalues			Extration Sums of Squared Loading			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.971	24.643	24.643	1.971	24.643	24.643	1.747	21.832	21.832
2	1.566	19.573	44.217	1.566	19.573	44.217	1.588	19.848	41.680
3	1.345	16.814	61.031	1.345	16.814	61.031	1.511	18.892	60.572
4	1.141	14.257	75.288	1.141	14.257	75.288	1.177	14.715	75.288
5	.760	9.495	84.782						
6	.631	7.884	92.666						
7	.346	4.328	96.994						
8	.241	3.006	100.000						

Extraction Method: Principal Component Analysis.

Table 6: Variance Explanation through Factor Analysis

Factor I

The 1st factor has an eigenvalue = 1.971. Since this is greater than 1.0, it explains more variance than a single variable, in fact 1.971 times as much.

The percent a variance explaines in this case is (1.971 / 8 units of variance) (100) = 24.643%

Factor II

The 2nd factor has an eigenvalue = 1.566. It is also greater than 1, and therefore explains more variance than a single variable. The percent a variance explains in this case is (1.566/ 8 units of variance) (100) = 19.573 %

Factor III

The 3rd factor has an eigenvalue = 1.354. Like Factors I & II it is greater than 1.0, and therefore explains more variance than a single variable. The percent a variance explains in this case is (1.354 / 8 units of variance) (100) = 16.814%

Factor IV

The 4th factor has an eigenvalue = 1.141. Like Factors I,II and III it is greater than 1.0, and therefore explains more variance than a single variable. The percent a variance explains in this case is (1.141 / 8 units of variance) (100) = 14.257.

The total cumulative variances defined by these four factors are 75.288 which means close to 76%.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
Zscore (Age)	-.019	-.056	.839	-.149
Zscore (Time)	.922	-.005	-.048	.036
Zscore (Tenure)	.927	.027	-.091	-.023
Zscore(Interest rate)	.128	.293	-.186	.651
Zscore (PCRI)	.005	.849	.247	.056
Zscore (Loan)	.009	.864	-.259	-.017
Zscore (CIBIL)	-.094	-.175	.092	.849
Zscore (Education)	-.111	.038	.791	.078

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Table 7: Rotated Component Matrix

High Load Factor In component 1

Time Lag

Tenure

High Load Factors in component 2

PCRI

Loan

High Load factor in component 3

Age

Education Qualification

High Load factor in component 4

Cibil

So, these are the critical factors.

Predictive Model for judging defaulters from non-defaulters.

For this process discriminant analysis is performed

Table of eigenvalues

It provides information on each of the discriminant functions (equations) produced. The study has been done using two groups, namely 'Default 'and 'no Default', so only one function is displayed.

Out of 81 sample size 41 are in the Default list and 40 are in Non- Default list. Defaulters are given the score as 1 (High credit risk) and Non-Defaulters are given score 2 (low Credit risk)

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.939[a]	100.0	100.0	.812

Table 8: Eigen value Table

The canonical correlation is the multiple correlations between the predictors and the discriminant function. With only one function it provided an index of overall model fit which is interpreted as being the proportion of variance explained (R^2). A canonical correlation of .812 suggests the model explains 66% of the variation in the grouping variable, i.e. whether a respondent Defaults or not.

Wilk's Lambda

Test of Function(s)	Wilk's Lambda	Chi-square	df	Sig.
1	.340	81.384	7	.000

Table 9: Wilki's Lambda

Wilks' lambda indicates the significance of the discriminant function. The table indicates a highly significant function (p <.001) and provides the proportion of total variability not explained, i.e. it is the converse of the squared canonical correlation. So we have 34% of the variation in the data unexplained

Canonical Discriminant Function Coefficients

	Function
	1
Age	.786
Time	.217
Tenure	.087
PCRI	-.018
Loan	-.475
CIBIL	-.674
Education	.488

Table 10: Standardized Canonical Discriminant Function Coefficients

The interpretation of the discriminant coefficients (or weights) is like that in multiple regressions. The table above provides an index of the importance of each predictor like the standardized regression coefficients (beta's) does in multiple regression. The sign indicates the direction of the relationship. Age is the strongest predictor whereas PCRI is the weakest predictor. CIBIL and Education Qualification are strong predictor.

Structure Matrix

	Function
	1
Age	.664
Education	.349
Loan	-.272
CIBIL	-.249

Tenure	.035
Time	.032
PCRI	-.031

Table 11: Structure Matrix.

The table above gives another method for demonstrating the relative significance of the indicators and it could be seen beneath. The structure framework table demonstrates the connections of every variable with each separate capacity. These Pearson coefficients are structure coefficients or discriminant loadings. They serve like component loadings in factor analysis. This table also showed that Age is the strongest predictor whereas PCRI is the weakest predictor. CIBIL, loan and education Qualification are strong predictor. Sign shows the direction of the relationship.

Canonical Discriminant Function Coefficients

	Function
	1
Age	.140
Time	.014
Tenure	.002
PCRI	.000
LOAN	.000
CIBIL	-.010
Education	.485
(Constant)	-.206

Unstandardized coefficients

Table 12: Canonical Discriminant Function.

The canonical discriminant function coefficient table

These unstandardized coefficients *(b)* are used to create the discriminant function (equation).

It operates just like a regression equation. In this case we have

D=0.140*age+0.014*Time+0.002*tenure-0.000004*Pcri-0.000001*Loan-0.010*cibil+0.485*Education qualification-0.206.

The discriminant function coefficients *b* or standardized form *beta* both indicated the Partial contribution of each variable to the discriminant function controlling for all other Variables in the equation. They could be used to assess each variables unique contribution to the discriminant function and therefore provide information on the relative importance of each variable.

Group centroids table

A further way of interpreting discriminant analysis results is to describe each group in terms of its Prediction, using the group means of the predictor variables. These group means are called centroids. These are displayed in the group centroids table.

Functions at Group Centroids

predictiondefault	Function 1
1.00	-1.358
2.00	1.392

Unstandardized canonical discriminant functions evaluated at group means

Table 13: Group Centroid Table

Determination of mean value

In prediction 1 no. of respondents are 41, prediction 2 number of variables are 40. Mean is calculated using the formula,
(-1.358*41+1.392*40)/81.This give Mean as 0.000025, so 0.000025 will serve as the cut off. The Discriminant Value to the left of Mean would show possible defaulter whereas to the right of mean would show a possible non-defaulter.

Classification Result[a,o]

	prediction		Predicted Group Membership 1.00	2.00	Total
Original	Count	1.00	38	2	40
		2.00	2	39	41
	%	1.00	95.0	5.0	100.0
		2.00	4.9	95.1	100.0
Cross-	Count	1.00	37	3	40

validated[b]		2.00	5	36	41
	%	1.00	92.5	7.5	100.0
		2.00	12.2	87.8	100.0

a. 95.1% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 90.1% of cross-validated grouped cases correctly classified

Table 14: Classification Results

The cross validated set of data is a more honest presentation of the power of the discriminant function than that provided by the original classifications and often produces a poorer outcome. The cross validation is often termed a 'jack-knife' classification, in that it successively classifies all cases but one to develop a discriminant function and then categorizes the case that is left out. This process is repeated with each case left out in turn. This cross validation produced a more reliable function. The classification results reveal that 95.0% of respondents are classified correctly into Defaulter and non-defaulter groups. This overall predictive accuracy of the discriminant function is called the 'hit ratio'. Defaulter and non-defaulter are predicted in the same accuracy in the study that is 90.1% which is on the higher side as it tends to 100%.

Segmenting the customer base

After Discriminant analysis, cluster analysis with the critical factors is performed to segment the customer base, so that the target customer could be identified. Cluster analysis is Performed with 81 cases and the they are groups that are relatively homogeneous within themselves and heterogeneous between each other, on the basis of the seven high loading factors which are Age, Time lag, Tenure, PCRI, Loan , CIBIL and Education qualification.

Stage 1: In the first stage Hierarchical clustering is used to find out the number of clusters. From the Dendogram and the Agglomeration schedule,the researchers concluded that the number of clusters are 4.(see Attachment)

Stage 2: K-mean Clustering

In this step 81 data are classified and clustered in to 4 segments. The Z score (mean=0 and standard deviation=1) of the critical factors derived from the factor analysis is used for k mean Clustering. Squared Euclidean distance is used for calculating distance.

Initial Cluster Centers

	Cluster			
	1	2	3	4

Zscore(Age)	1.29002	2.21355	-2.14025	-.02932
Zscore(time)	-.93796	1.71849	1.32002	-.80513
Zscore(tenure)	-.97390	1.95998	1.37826	.11366
Zscore(interest)	-2.06340	-.09710	.55833	1.21377
Zscore(pcri)	-1.00999	2.87548	.26987	3.17661
Zscore(loan)	-.53418	-1.01098	3.94515	2.87797
Zscore(cibil)	-1.71792	-.01293	-2.28178	.53750
Zscore(gender)	1.26214	-.78253	-.78253	1.26214
Zscore(edu)	-.05558	.84476	-.05558	1.74510

Table 15: Initial cluster centre

The first step in k-means clustering is finding the k centers. This is done iteratively. The initial cluster centers are the variable value of the well spaced observation. After the initial cluster centers have been selected, each case is assigned to the closest cluster, based on its distance from the cluster centers. After all of the cases have been assigned to clusters, the cluster centers are recomputed, based on all of the cases in the cluster

Final Cluster Centers

	Cluster			
	1	2	3	4
Zscore(Age)	.05953	.70951	-.77694	-.06890
Zscore(time)	-.47579	1.33331	1.22041	-.46644
Zscore(tenure)	-.45412	.95083	1.45203	-.46806
Zscore(pcri)	-.35493	.29681	-.36447	1.87971
Zscore(loan)	-.37976	-.36677	.59037	1.51916
Zscore(cibil)	-.03348	.56972	-.60475	.32002
Zscore(edu)	.10979	.21452	-.50575	-.14561

Table 16: Final cluster centres

The Final cluster centre's are computed as the mean for each variable within each final cluster. The Final cluster centre reflects the characteristics of the typical case for each cluster.

1		2.214	2.955	3.536
2	2.214		2.889	3.761
3	2.955	2.889		3.812
4	3.536	3.761	3.812	

Table 17: Distance between Final cluster centers

This table shows the Euclidean distances between the final cluster centers. Greater distances between cluster corresponds to greater dissimilarities. Clusters 3 and cluster 4 are most different. Cluster 3 is approximately similar to cluster 1 and cluster 2.

Number of Cases in each

Cluster	1	49.000
	2	10.000
	3	12.000
	4	10.000
Valid		81.000
Missing		.000

Table 18: Number of cases in each cluster

It is clear that 49 percent of the loan getters lie in cluster 1, 10 percent in cluster 2, 12percent in cluster 3 and 10 percent in cluster 4.

The final cluster centre describes the mean value of each variable for each of the 4 clusters.

Cluster 1

Based on the mean of the z value, it could be observed that this cluster is dominated by moderate age, low time lag, low tenure, low PCRI, low loan, low CIBIL and high education qualification.

Cluster 2

Based on the mean of the z value, it could be observed that this cluster is dominated by High age, high time lag, moderate tenure, high PCRI, low loan, high CIBIL scores with moderate education.

Cluster 3

Based on the mean of the z value, it can be observed that this cluster is dominated by low age, high time lag, high tenure, low PCRI, moderate loan, low CIBIL and low education qualification

Cluster 4

Based on the mean of the z value, it can be observed that this cluster is dominated by low age, low time lag, low tenure, high PCRI, high loan, moderate CIBIL and low education qualification.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(Age)	4.166	3	.877	77	4.753	.004
Zscore(time)	16.306	3	.404	77	40.395	.000
score(tenure)	15.546	3	.433	77	35.880	.000
Zscore(pcri)	14.660	3	.468	77	31.340	.000
Zscore(loan)	11.891	3	.576	77	20.656	.000
Zscore(edu)	1.444	3	.983	77	1.470	.229
Zscore(cibil)	2.905	3	.926	77	3.137	.030

Table 19: Dispersion Analysis

From the dispersion analysis it is clear that time lag had the greatest influence in forming the cluster followed by tenure and PCRI while CIBIL score had the lowest influence in forming the cluster.

CONCLUSION

Education qualification of the borrower shows a strong degree of association with the loan default whereas gender, number of dependents, and source of income found to have no association with the Loan Default. Factor analysis is performed which showed Age, Time lag, Tenure, PCRI, CIBIL score, Loan amount and education qualification as the critical factors and based on those factors discriminant investigation is conducted to anticipate whether a customer is a possible defaulter or a non-defaulter with Indicator variables such as age, time lag, tenure, PCRI ,Loan amount ,Education qualification and CIBIL score. These variables are deduced from the factor analysis as the high loading factors. The discriminant function revealed a significant association between groups and all predictors, accounting for 66% of between group variability, although closer analysis of the structure matrix revealed four significant predictors, namely age, loan, CIBIL and education, in which CIBIL and loan shows a negative relationship. Time lag and PCRI are considered to be weakest predictor. The cross validated classification shows that overall 90.1% are correctly classified. A cluster analysis is run on 81 cases, each responding to Variables on age, time lag, tenure, interest, PCRI, loan and CIBIL.A hierarchical cluster analysis produces four clusters, between which the variables are significantly different in the main. The first cluster is dominated by moderate age, low time lag, low tenure, low PCRI, low loan, low CIBIL and high education qualification. The second cluster is dominated by high age, high time lag, moderate tenure, high PCRI, low loan, high CIBIL scores with moderate education. The third cluster is dominated by low age, high time lag, high tenure, low PCRI, moderate loan, low CIBIL and low education qualification. The fourth cluster is dominated by low age, low time lag, low tenure, high PCRI, high loan, moderate CIBIL and low education qualification. The CIBIL score is not given due importance by this financial institute

Limitations of the study

The study is carried out in a particular city (Kolkata). The sample size is statistically significant but small in size. The researcher wanted to study the data but due to time constrain it is not possible. Some of the information about the customer personal information like geographic location, contact information is kept confidential so the geographic segmentation and geographic location is difficult to trace.

Future Scope

The study should be carried out with a large sample size including data from different geographic location. Positioning of the firm should be done with respect to its nearest competitors.

References

1. *Ahmad, S. A. (1997). Natural Hazards and Hazard Management in the Greater Caribbean and Latin America. 3.*

2. *Akinwumi, J. A., & Ajayi, A. O. (1990). The Role of Loan Repayment Capacity in Small Scale Farmers Credit Administration. Quarterly Journal of Credit Administration, 15(2).*

3. *Ameyaw, A. I. (2011). Causes and effects of loan defaults on the profitability of Okomfo Anokye Rural Bank.*

4. *Bennett, P., Peach, R., & Peristiani, S. (1997). Structural change in the mortgage market and the propensity to refinance. Federal Reserve Bank of network(9736).*

5. *Berger, A. N., & De , Y. R. (1995). Problem loans and cost efficiency in commercial Banks. Economic and Policy Analysis paper.*

6. Balogun, E. a. (1998). Loan Delinquency Among Small Farmers in Developing Countries: A Case Study of the Small- Farmers Credit Programme in Lagos State in Nigeria. *CBN Economic and Financial Review* (3), 26.

7. Balogun, E. D., & Alimi, A. (1990). Loan Delinquency Among Small Farmers in Developing Countries: A case study of the Small- Farmers Credit Programme in Lagos State of Nigeria. *CBN Economic and Financial Review* (3), 26.

8. Campbell, T., & Dietrich, J. K. (1983). The Determinants of Default on insured conventional Residential Mortgage Loans. *The Journal of Finance , 38* (5), 1569-1581.

9. *Cairney, J., & Boyle, M. H. (2004). Home ownership, mortgages and psychological distress. Housing Studies, 19(2), 161-174.*

10. *Capozza, D. R., Kazarian, D., & Thomson, T. A. (1997). Mortgage default in local markets. Real Estate Economics, 25(4), 631– 655.*

11. *Hakim, R. S., & Haddad, M. (1999). Borrower attributes and the risk of default of conventional mortgage. Atlantic Economic Journal, 27(2), 210–220.*

12. Har, N. P., & Eng, O. S. (2004). Risk sharing in mortgage loan agreements. *Review of Pacific Basin Financial Markets and Policies , 7* (2), 233–258.

13. *Itoo , R. A., Mutharasu, S. A., & Filipe, J. A. (2013). Effect of Loan Value and Collateral on Value of Mortgage Default. International Journal of Latest Trend in Finance and Economic Sciences, 3(4), 635-651.*

14. *Jacobson, T., & Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. Journal of Banking & Finance, 27(4), 615-633.*

15. *Jackson, R. J., & Kaserman, L. D. (1980). Default risk on Home mortgage loans: A test of competing Hypothesis. The Journal of Risk and Insurance, 47(4), 678-690.*

16. *Korankye , A. A. (2014). Causes and Control of Loan Default/Delinquency in Microfinance Institutions in Ghana. American International Journal of Contemporary Research, 4(12), 36-45.*

17. *Kumar, M. (2010). Demographic Profile As a Determinant of Default Risk in Housing Loan Borrowers. Applicable to Indian Condition, International Research Journal of Finance and Economics, 58,1450-2887.*

18. *Liu, D. Y., & Lee, S. P. (1997). An analysis of risk classifications for residential mortgage loans. Journal of Property Finance, 8(3), 207-225.*

19. *Loan Delinquency Among Small Farmers in Developing Countries: A Case Study of the Small-Farmer Credit Programme in Lagos State of Nigeria, 26(3) (1990).*

20. M.J., G. A. (2003, March). Can Credit Registries Reduce credit Constraints? Empirical Evidence on the Role of Credit Registries in Firm Investment Decisions. *Annual Meetings of the Inter- American Development Bank* . Santiago, Chile.

21. *Murray, J. (2011). Default on a loan. United States Business Law and Taxes Guide*

22. *Okorie, A. (1986). Major Determinants of Agricultural Loan Repayments, Savings and Development.*

23. *Olomola, A. S. (1999). The nature and determinants of rural loan repayment performance in Nigeria: The case of FADU's micro-credit programme. NISER monograph series NO Ibadan.*

24. *Pearson , R., & Greef, M. (2006). Causes of Default among Housing Micro Loan Clients, FinMark Trust Rural Housing Loan Fund, National Housing Finance Corporation and Development Bank of Southern Africa.*

25. Teo, A. H., & Ong, S. E. (2005). Conditional default risk in housing arms: A bivariate probit approach. *13th American Real Estate Society Conference, Santa Fe, USA.*

Average Linkage (Between Groups)

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	49	50	.579	0	0	26
2	44	68	.613	0	0	14
3	22	29	.698	0	0	9
4	9	72	.792	0	0	9
5	71	73	.879	0	0	54
6	10	30	1.035	0	0	32
7	46	48	1.099	0	0	12
8	45	55	1.294	0	0	21
9	9	22	1.519	4	3	38
10	16	42	1.534	0	0	27
11	77	80	1.544	0	0	62
12	46	69	1.569	7	0	31
13	4	21	1.573	0	0	33
14	40	44	1.667	0	2	23
15	32	39	1.681	0	0	38
16	2	15	1.739	0	0	66
17	11	14	1.835	0	0	53
18	28	38	1.835	0	0	32
19	54	57	1.863	0	0	29
20	19	31	1.863	0	0	67
21	45	66	1.927	8	0	48
22	51	74	1.930	0	0	39
23	40	53	1.935	14	0	34
24	13	17	2.052	0	0	42

25	18	34	2.077	0	0	28
26	26	49	2.537	0	1	37
27	12	16	2.573	0	10	42
28	18	33	2.609	25	0	56
29	54	59	2.719	19	0	49
30	43	64	2.818	0	0	35
31	46	56	2.885	12	0	37
32	10	28	2.918	6	18	51
33	4	27	2.924	13	0	61
34	40	62	2.987	23	0	41
35	43	47	2.992	30	0	47
36	61	79	3.032	0	0	54
37	26	46	3.201	26	31	48
38	9	32	3.296	9	15	41
39	51	58	3.378	22	0	50
40	3	20	3.390	0	0	63
41	9	40	3.514	38	34	53
42	12	13	3.591	27	24	43
43	12	37	3.618	42	0	51
44	65	67	3.748	0	0	47
45	5	23	3.805	0	0	55
46	35	63	3.953	0	0	65
47	43	65	4.396	35	44	60
48	26	45	4.492	37	21	56
49	54	60	4.532	29	0	61
50	51	52	4.533	39	0	66
51	10	12	4.576	32	43	58
52	76	81	4.616	0	0	62
53	9	11	5.366	41	17	59
54	61	71	5.431	36	5	60

55	5	36	5.597	45	0	64
56	18	26	5.708	28	48	59
57	6	7	6.167	0	0	70
58	10	70	6.285	51	0	63
59	9	18	6.365	53	56	68
60	43	61	7.091	47	54	64
61	4	54	7.133	33	49	71
62	76	77	7.405	52	11	65
63	3	10	7.743	40	58	68
64	5	43	8.064	55	60	73
65	35	76	8.193	46	62	74
66	2	51	8.255	16	50	71
67	19	25	8.375	20	0	70
68	3	9	9.086	63	59	72
69	1	8	9.497	0	0	72
70	6	19	10.174	57	67	76
71	2	4	10.485	66	61	77
72	1	3	10.888	69	68	73
73	1	5	11.508	72	64	75
74	35	75	12.556	65	0	75
75	1	35	13.491	73	74	77
76	6	41	14.749	70	0	78
77	1	2	15.494	75	71	78
78	1	6	22.777	77	76	79
79	1	78	28.618	78	0	80
80	1	24	36.433	79	0	0

Attachment 1: Agglomeration schedule

Cluster Membership

Case Number	Cluster	Distance
1	1	3.020
2	3	1.785
3	3	1.867
4	3	2.390
5	1	2.424
6	4	2.398
7	4	2.280
8	1	3.426
9	1	1.551
10	1	2.213
11	4	2.143
12	1	2.054
13	3	2.780
14	1	2.648
15	3	2.300
16	2	2.222
17	3	1.761
18	1	2.424
19	1	3.904
20	1	3.088
21	4	2.581
22	1	2.245
23	1	2.733
24	3	4.318
25	4	2.354
26	2	2.109
27	3	1.927
28	1	2.809
29	1	1.396
30	1	2.380
31	4	2.684
32	1	1.951
33	1	1.866
34	1	2.906
35	1	3.468
36	1	2.517
37	2	2.255
38	1	2.549
39	1	1.878
40	1	1.629
41	4	2.809
42	1	3.197
43	2	2.179
44	1	1.279
45	1	2.043
46	2	1.734
47	1	2.607
48	1	2.389
49	1	1.964
50	1	1.846
51	3	1.500
52	3	2.352
53	2	2.295
54	2	2.268
55	1	2.737
56	2	1.685
57	2	2.446

58	3	2.132
59	2	1.968
60	2	2.927
61	1	3.175
62	1	2.239
63	1	2.332
64	1	3.324
65	2	2.851
66	1	3.250
67	1	3.121
68	1	2.380
69	2	1.746
70	3	2.386
71	2	1.995
72	1	1.517
73	1	2.482
74	3	3.204
75	2	2.325
76	2	2.862
77	2	1.936
78	2	3.786
79	2	2.350
80	2	2.485
81	2	2.304

Attachment 2: Final cluster centers.