

An empirical study to identify significant factors and their influence on credit worthiness of first time loan buyers

Dr. Jyotsna Munshi

MET League of Colleges, Mumbai, India

ABSTRACT: We are living in an era, where the entire world is reeling under the influence of digitization. This digitization has generated huge amount of data from various systems. Data is commonly referred as the new oil and as a new commodity which eventually helps in improvement of products and services, satisfying customers, maximizing profitability, and building of efficient operational systems. The data can be analyzed to derive meaningful insights and make predictions which will help in adding value to the business and making them more profitable. Credit Lending companies want to modify their lending policies by offering money to the customers who have no or not sufficient credit histories. They want to achieve this by maintaining the adjustable risk limits and ensure safe borrowing experience. The probability of default always remains as the area of concern for all credit lending companies. Companies use the data obtained from different channels to determine the probability of default for a prospective and a potential borrower. A data of 15,375 customers was chosen and analyzed using different tools and techniques. Different significant factors and their influence on credit worthiness of first-time loan buyers with no or insufficient credit history were studied and evaluated. Based on the significance level of the factors, different models were predicted based on different techniques. A table comparing error rates of different models was also prepared in order to offer quick reference to the lending companies for the selection of an appropriate model for their implementation purpose.

KEYWORDS—Credit, Default, Lending, Loan, Repayment

I. INTRODUCTION

Much water has flown down the bridge since early man appeared in this world. With the continuous growth of the civilization, the needs of human race are being fulfilled in the best possible way. While all the needs are fulfilled, there also exists a bucket list of wants. This gives rise to need of additional capital and then the role of money lending agencies comes into play. The lending market in India has grown considerably over a period of time. The lending market in India grew to Rs 174.3 lakh crore in March 2022, up by 11.1 per cent on an annual basis, as compared to March 2021 (Outlook, 2022).

The Finance and Banking sector is facing many problems in lending credit and loans to the borrowers. Normally, companies rely majorly on credit histories to evaluate the risk associated with borrowers. Consequently, the individuals who don't have or have insufficient credit records are unable to make use of the loan facility. People struggle to get loans due to insufficient or non-existent credit histories. It becomes imperative for Credit lending companies to give a scope for the unbanked population by providing a positive and safe borrowing experience. Institutions are required to adapt a set of practices to cater to a large customer base, while keeping track of adjustable risk limit and assuring a safe borrowing experience.

The study focused on how credit lending companies need to modify their lending policies to extend the loan offers to individuals with minimal or no credit history. The project discusses about identifying the important parameters and developing a framework that will help companies for assessing the creditworthiness of customers and minimizing the probability of default.

The primary objective of the study was to identify the significant factors for the credit lending companies in order to predict the decision to advance or not to advance a loan. This would help the lending companies to mitigate the risk associated by lending to individuals with limited credit history. This would eventually help them in enhancing customer base. This would help the customers with insufficient or no credit history to avail

loan facility.

The study discusses the specific objectives, methodologies, and analysis outcomes of the study. The study is useful for lenders as well as borrowers.

II. MOTIVATION

The probability of default always remains as the area of concern for all credit lending companies. Companies use the data obtained from different channels to determine the probability of default for a prospective and a potential borrower. Through this project, the attempt has been made to demonstrate the application of business analytics to redefine and redesign the lending practices. This would help credit lending companies in minimizing risk and adapting to changing customer needs. This would also help customers with no or minimal credit history to avail the loan facility.

III. RESEARCH PROBLEM IDENTIFICATION

The Credit Lending companies do face a tough crucial challenge while extending loans to individuals with no or insufficient credit histories to ensure a safe borrowing experience. They need a simple but potent and analytics driven solution to mitigate their risks while ensuring a safe borrowing experience.

IV. LITERATURE REVIEW

Model for credit worthiness was created.(Bai, Shi, Liu, & Sarkis, 2019).Modelinvolved combining of Fuzzy rough set and Fuzzy C-means clustering. • The model was tested using actual bank data of 2044 farmers of China and was used to predict creditworthiness of farmers. • Education and skills werethe two most important factors which enhance credit-worthiness of farmers.

(Gharat, 2020) discussed about the transformation in BFSI sector in India in mid 90s with the private and multinational banks coming in to the picture. Due to enormous economic reforms in the country, there was increase in the number of banks. With more number of banks and the hustle in the banking sector, the banking activities increased manifold and affected many areas of operation of banks, particularly in the field of retail lending. Modus operandi opted by Banks remained in the terms of giving credit against security given by its customers associated with the bank. Credit rating determines that how much a lending agency can trust a certain loan buyer, whether an individual, a corporation, or a country. The credit rating is predicted using borrowers past history and the existing financials.

(Gorgijevska & Gjorgieva-Trajkovska, 2019)analyzed the overallfinancial stability of the client and the specific project before financing the loan. This is done to have the assessment of the credit risk or the credit worthiness of the loan applicant. The mainobjective of this activity is evaluate the capacity of repayment of loans.The banks use several methods for the qualitative analysis like CAMPARI, PEST, SWOT, etc.

(Jabocci, 2009)discusses how due to recession, the situation has become like olden times where retail lending was decided on the credit history and the ability to repay. The article discussed that the most important step towards getting a loan for a business house was by having a sound finance with a good relation with agent. It was emphasized that Businesses should have strong financial knowledge. Businesses should present impressively that how the loan amount will be used and how the fluctuations in sales will be handled and projections kept in order.

(Joel, 2014)discussedabout how the various political connections affect the cost and terms of loan contracts. Data was collected from the political connections of around 500 companies.It was observed that the cost of bank loans was lower for companies that have strong political ties. This was explained in two ways:One in which retail lenders charge lower rates because they are confident about the borrower's credit worthiness and second in which a banks offerhigher value to connected loans to strengthen their relationships with top political big wigs.

(Kevin, 2017) discussed, how the text analysis was used in evaluating credit worthiness for loan applications. This included the identifying the different traits of applicants with their choice and use of words and phrases which are not specified in credit reports, and risks of discriminatory lenders charges involving default or commonly used for a specific ethnic group. It was observed that the probability of repayment is more if the communication clear and crisp.

The lending market in India has grown considerably over a period of time. The lending market in India grew to Rs 174.3 lakh crore in March 2022, up by 11.1 per cent on an annual basis, as compared to March 2021 (Outlook, 2022).

(Sheila, Inhyuck, Jose-Luis, & Myers, 2015) studied differences in loan denial rates for different racial/ethnic groups. The gaps were explained in two different ways: one was that the disparity resulted from underlying

racial disparities in credit worthiness and another view was that the disparities arise because of a pattern of racial discrimination among retail lenders. The study used economic stratification approach to evaluate the assumptions. The results indicated that one of the reasons of poor credit risk among black applicants is that blacks with good credit risk undervalue their credit worthiness and number of loan applications are low. Findings suggested that even in the case of nondiscriminatory lending behavior, there is a high probability of screening out low-risk blacks and eventually leading to higher denial rates among blacks.

(Statista, 2022) Lending and borrowing money are crucial for the financial health of a country. Traditionally, banks were the primary lenders in India for disbursing loans (Commercial and personal). Cumbersome processes followed by banks led to individuals shifting to the unorganized sector or traditional moneylenders with exorbitant interest rates. However, in the recent years, the lending market in India has developed into a dynamic playfield of various non-banking financial companies, fintech enterprises, and digital lending platforms. More often, Indian banks faced the issue of non-performing assets (NPA). The government motivated banks to modify their lending policies to overcome the crisis. Banks switched from corporate lending to the personal loan category to reduce the risk of default. A large number of small personal loans were offered over big fat loans. India's is large. A salary cut in recent times and increased costs because of inflation requires the financing of personal requirements through loans. Additional expenditure in the festive seasons further leads to demand for personal loans. The consumer lending market is changing and improving because of digital lending platforms and online credit options. RBI aims to keep a check on the growth of unregulated digital lending players for seamless compliance with data and privacy protection. RBI has issued guidelines for various reforms. This new age transformation will ensure consumer confidence and trust in digital lending platforms in India. This will eventually lead to a well operated and safe digital lending ecosystem for the country.

V. DATA DESCRIPTION

Data was sourced from (Kaggle.com, 2017). A competition on Home Credit Default Risk was posted on Kaggle 5 years ago. A data of 15,375 customers was chosen from the data base for the project.

Data of Customers: 15, 375

Number of Variables: 14

Data Labels: Gender, Marital status, Children, Own car, Own house, Service status, Family income, Type of loan, Loan Application day, Loan application time, Loan Amount credited, Actual loan price item, Region Rating, Default.

Dependent Variable: Default

Independent Variable: Gender, Marital status, Children, Own car, Own house, Service status, Family income, Type of loan, Loan Application day, Loan application time, Loan Amount credited, Actual loan price item, Region Rating.

Dataset Description:

1. Gender: Gender had two labels: Males and Females. There are 66.06% females and 33.94% males. Refer Fig 1.
2. Marital Status: Marital Status had three labels: Married, Single/Separated and Widow. There were 73.10% married, 21.39% from single/separated and 5.51% widow. Refer Fig 2.
3. Children: Children had 12 labels. 69.79% had no children. 20.34% have 1 child. 8.55% have 2 children and remaining 1.16% contribute to the remaining labels. Refer Fig 3.
4. Own Car: Own car had 2 labels. 65.74% of the sample set had no car and 34.26% own a car. Refer Fig 4 in Annexure.
5. Own House: Own House had 2 labels. 68.68% own car and 31.32% do not own a car. Refer Fig 5.
6. Service Status: Service Status had 5 labels. 73.35% were into private service. 7.04% were into Govt service. 17.59% were pensioners, 0.01% were into business and 0.01% were students. Refer Fig 6.
7. Family Income: Family income had 10 labels. 46.35% had income in the range of (1L-2L per month), 28.92% had income in the range of (2L-3L per month), 9.66% had income in the range of (25K-1L per month), 8.82% had income in the range of (3L-4L per month), 3.65% had income in the range of (4L-5L per month) and rest contributed to the remaining population. Refer Fig 7.
8. Type of Loan: Type of loan had 2 labels. 90.28% had opted for cash loans and 9.72% opted for

revolving loans. Refer Fig 8.

9. Application Day: Application day had 7 labels. 18.02% had started loan on Tuesday, 16.93% had started loan on Wednesday, 16.57% on Thursday, 16.06% on Monday, 15.64% on Friday, 11.19% on Saturday and 5.59% on Sunday. Refer Fig 9.
10. Application Time: Application time had 24 labels.
11. Loan Amount credited: Loan amount credited had 9 labels. 39.32% availed loan in the range of (5L-10L), 34.96% had availed (1L-5L), 18.49% availed (10L-15L) and rest contributed to the remaining percentage. Refer (Figure 10 in Annexure).
12. Actual Loan Price item: Actual loan price had 9 labels.
13. Region Rating: Region rating had 3 labels. Region 1 contributed 5.08%, Region 2 contributed 72.01% and Region 3 contributed 22.91%. Refer Fig 11.
14. Default: Default had 2 labels. 91.59% were not Defaulters. 8.41% were defaulters. Refer Fig 12.

VI. METHODOLOGY

Data Collection

Data was sourced from (Kaggle.com, 2017). A competition on Home Credit Default Risk was posted on Kaggle posted 5 years ago. A huge data set was available. Data was cleaned and specific variables were selected to predict the default rating. Data, where 91.59% were successful in repayment of loans was chosen, in order to identify the correct variables to ensure safe lending practices for the credit lending companies. A data of 15,375 customers was chosen from the data base for the project.

Business Questions

Using exploratory data analyses following issues were addressed:

- Understanding the repayment and default patterns across different customers.
- Identifying the important factors useful for credit lending companies for safe lending experience.
- Identifying the correlation between different factors and the default patterns across different customers.
- Predicting a suitable model for the credit lending companies for safe lending and ensuring minimum risks.

Hypotheses developed:

- H01: There is no significant relation between the marital status and the default patterns across different customers.
- H02: There is no significant relation between the service status and the default patterns across different customers.
- H03: There is no significant relation between the family income and the default patterns across different customers.
- H04: There is no significant relation between loan amount credited and the default patterns across different customers.
- H05: There is no significant relation between region rating and the default patterns across different categories.

Tools and Techniques Used:

- Data exploration of all the data variables using MS-Excel.
- Correlation with Cross Tab using SPSS.
- Hypotheses testing with ANOVA using SPSS.
- Logistics Regression using SPSS.
- Machine Learning (Tree, Bagging and Random Forest) using R-Software.
-

VII. DATA ANALYSIS

CORRELATION USING CROSSTAB

- 91.9% of population was successful in repayment of loans and 8.1% were defaulters. For females, 93.2% were successful in repayment of loans and 6.8% were defaulters. For males, 89.5% were successful in repayment of loans and 10.5% were defaulters. (Refer Table 1(a)).
- Gender and Default were observed to be highly correlated. The significant factor was 0.000. (Refer

Table 1(b)).

- For widows, 95.4% were successful in repayment of loans and 4.6% were defaulters. For single/separated, 92% were successful in repayment of loans and 8% were defaulters. For married, 90.7% were successful in repayment of loans and 9.3% were defaulters. (Refer Table 2(a)).
- Marital Status and default were observed to be highly correlated. The significant factor was 0.000. (Refer Table 2(b)).
- For population with no children, 92.2% were successful in repayment of loans and 9% were defaulters. (Refer Table 3(a)).
- There was no correlation between children and default. (Refer Table 3(b)).
- 92.2% having car were successful in repayment of loans and 7.8% were defaulters. (Refer Table 4(a)).
- There was no correlation between owning a car and default. (Refer Table 4(b)).
- 92.1% having own house were successful in repayment of loans and 7.9% were defaulters. (Refer Table 5(a)).
- There was no correlation between owning a house and default. (Refer Table 5(b)).
- 91.2% from were successful in repayment of loans and 8.8% were defaulters. (Refer Table 6(a)).
- Service Status and default were observed to be highly correlated. The significant factor was 0.000. (Refer Table 6(b)).
- There was no correlation between family income and default. (Refer Table 7(b)).
- There was no correlation between type of loan and default. (Refer Table 8(b)).
- There was no correlation between loan application day and default. (Refer Table 9(b)).
- Loan application time and default were observed to be highly correlated. The significant factor was 0.005. (Refer Table 10(b)).
- There was no correlation between loan amount credited and default. (Refer Table 11(b)).
- Actual loan price and default were observed to be highly correlated. The significant factor was 0.005. (Refer Table 12(b)).
- Actual rating price and default were observed to be highly correlated. The significant factor was 0.005. (Refer Table 13(b)).

HYPOTHESES TESTING USING ANOVA

- It was observed that there was significant relation between marital status and Default. The significant factor was 0.0000. Hence, null hypothesis was rejected. (Refer Table 14).
- It was observed that there was significant relation between service status and Default. The significant factor was 0.0000. Hence, null hypothesis was rejected. (Refer Table 15).
- It was observed that there was no significant relation between family income and Default. The significant factor was 0.481. Hence, null hypothesis was accepted. (Refer Table 16).
- It was observed that there was significant relation between loan amount credited and Default. The significant factor was 0.003. Hence, null hypothesis was rejected. (Refer Table 17).
- It was observed that there was significant relation between region rating and Default. The significant factor was 0.0000. Hence, null hypothesis was rejected. (Refer Table 18).

FACTOR ANALYSIS

- The value of KMO test is 0.584. The value was more than 0.5 and close to 0.6; hence factor analysis was done. (Refer Table 19).
- The Bartlett's test indicates significance value of 0.000. This indicated that correlation matrix was an identity matrix and there was no correlation between the variables. (Refer Table 19).
- 6 factors were extracted explaining 63.11% of variance. (Refer Table 20)
- Factors were extracted using Principal Component Analysis. (Refer Table 21 and 22).
- Based on the results from Principal Component Analysis, variable loan application day was dropped.
- KMO test was redone. The value of KMO test was 0.584. The value was more than 0.5 and close to 0.6; hence factor analysis was done. (Refer Table 24).
- The Bartlett's test was redone. The Bartlett's test indicates significance value of 0.000. This indicated that correlation matrix was an identity matrix and there was no correlation between the variables. (Refer Table 24).
- 6 factors were extracted explaining 68.759% of variance. (Refer Table 25)
- Factors were extracted using Principal Component Analysis. (Refer Table 26 and 27).

- Factor scores were analyzed and it was observed that they were not correlated. (Refer Table 29).
- Factor scores were analyzed and descriptives were computed. It was observed that Mean was 0 and Standard Deviation was 1. (Refer Table 30).

LOGISTICS REGRESSION

- Logistics Regression was done to predict the probability of Default and to identify important parameters affecting Default.
- Model was significant with value 0.0000. (Refer Table 32).
- Variables included in the equation were: Gender, Marital status, Service status, Type of loan, Loan application time, Actual loan price and Region rating. (Refer Table 33).
- Model was validated. Model was significant with value 0.0000. (Refer Table 34).
- The value for Hosmer and Lemeshow test was 0.673, indicating a good fit. (Refer Table 35).
- Confusion Matrix was predicted and the error rate was calculated as 8%. (Refer Table 36). The cut off probability for the model was 0.5. (Refer Table 37)
- Variables included in the final model were: Gender, Marital status, Own Car, Service status, Type of loan, Loan application time, Actual loan price and Region rating. (Refer Table 38).

MACHINE LEARNING TECHNIQUES

TREE

- Initial tree was grown. (Reference Annexure Page No.). 14 independent nodes were observed.
- Error rate was calculated as 12.4 %.
- Best fit was observed as nodes=7. (Reference, Fig 13).
- New tree was grown with (Best=7).
- Error rate was calculated as 11.2 %.

BAGGING

- 500 trees were grown, with 12 variables at each split.
- Error rate was calculated as 8.33 %.

RANDOM FOREST

- 500 trees were grown, with 4 random variables at each split. (Reference Annexure Page No.).
- Error rate was calculated as 8.1 %.
- Importance of variables was calculated. It was observed that Loan amount credited was the most important variable, followed by Family income, Loan application time and Loan application day.
- Variable importance plot was plotted. (Reference Fig14).

VIII. CONCLUSION

- Based on the Cross-Tabs, variables (Gender, Marital Status, Service Status, Loan Application time, Actual loan price and Region rating) were correlated and significant.
- Based on Logistics regression, variables (Gender, Marital status, Own Car, Service status, Type of loan, Loan application time, Actual loan price and Region rating) were significant and model was predicted.
- Based on Tree, variables (Gender, Loan application day, Actual loan price and children) were significant and model was predicted.
- Based on Random Forest (Loan amount credited, Family income per month, Loan application time and Loan application day) were significant.
- Error rate calculated observed that the error rate was almost similar in Logistics regression and Random forest, followed closely by Bagging and for the Tree configuration, the error rate was high by around 34%. (Reference Table No.39).
- There were more females than men. However, the percentage default for men was more than women. (Reference Table 1(a)).
- Individuals who were Single/separated contributed to 72.1 % default compared to ones who own a car. (Reference Table 2(a)).
- Families having no children contributed to 67.3% of default compared to other categories. (Reference Table 3(a)).
- Individuals who do not own a car contribute to 67% default compared to ones who own a car.

(Reference Table 4(a)).

- Individuals who do not own a house contributed to 67% default compared to ones who do not own a house. (Reference Table 5(a)).
- Individuals in private service contributed to 82.5% default compared to other categories. (Reference Table 6(a)).
- Individuals with Family income (1L-2L), contributed to 49.7% default compared to other categories. (Reference Table 7(a)).
- Cash loans contributed to 94.2 % default compared to other categories. (Reference Table 8(a)).
- Wednesdays contributed to 18.7 % default compared to other categories. (Reference Table 9(a)).
- Loans given at 12Noon contributed to 13.3 % default compared to other categories. (Reference Table 10(a)).
- Loan credited in the range of (1L-5L), contributed to 47.6% default compared to other categories. (Reference Table 11(a)).
- Actual loan price in the range of (1L-5L), contributed to 63.1% default compared to other categories. (Reference Table 12(a)).
- Region rating 2 contributed to 73.3% default compared to other categories. (Reference Table 13(a)).

IX. RECOMMENDATIONS

- Age group of the customers should be mentioned. This will help credit lending companies in understanding the effective time customers have for the loan repayment.
- Educational qualification of the customers should be mentioned. This will help credit lending companies to gauge the awareness of customers about the basic understanding of loans and their repayment schedules.
- Type of loans should be mentioned. It is imperative to understand whether loan is for home, vehicle or for education. This will help in understanding which loans are in demand and which are the ones mostly defaulted.
- Tenure for Loan repayment should be mentioned. This will help in understanding comfort level of customers for the tenure and consequently the repayment.

REFERENCES

- [1.] Bai, C., Shi, B., Liu, F., & Sarkis, J. (2019, March). Banking credit worthiness: Evaluating the complex relationships. *Science Direct.com*, 83, 26-38.
- [2.] Gharat, P. U. (2020). *A study on credit rating agencies of India*. Thakur College of Science and Commerce, Mumbai.
- [3.] Gorgijevska, A., & Gjorgieva-Trajkovska, O. (2019). Qualitative and quantitative analysis of creditworthiness of the companies. *Journal of Economics*, 4(1), 18-26.
- [4.] Jabocci, M. S. (2009, May 18). *Business need to prove credit-worthiness*. Retrieved from www.westfaironline.com.
- [5.] Joel, H. F. (2014, March). Political Connections and the Cost of Bank Loans. *Journal of Accounting Research*, 52(1), 193-243.
- [6.] Kaggle.com. (2017). *Kaggle*. Retrieved April 2023 3, 2023, from <https://www.kaggle.com/c/home-credit-default-risk>: <https://www.kaggle.com/c/home-credit-default-risk>
- [7.] Kevin, W. (2017, January 11). Can words used in loan applications reveal creditworthiness? *American Banker*, 182(210).
- [8.] Outlook. (2022, September 20th). *Outlook Money*. Retrieved June 8th, 2023, from Outlook Money: <https://www.outlookindia.com/business/india-s-lending-market-grew-11-1-annually-in-fy-2022-says-credit-bureau-report-news-224599>
- [9.] Sheila, A., Inhyuck, H., Jose-Luis, M., & Myers, S. (2015, June). Bad credit and Intergroup differences in Loan Denial Rates. *Sage Journals*, 42(1-2), 19-34.
- [10.] Statista. (2022, December 19th). *Lending market in India - statistics and facts*. Retrieved June 8th, 2023, from Statista: <https://www.statista.com/topics/10214/lending-market-in-india/#topicOverview>

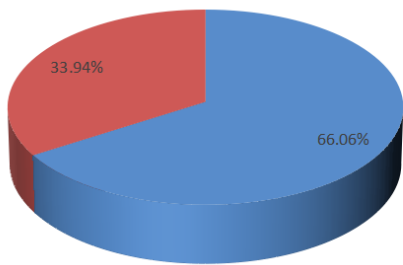


Figure 1

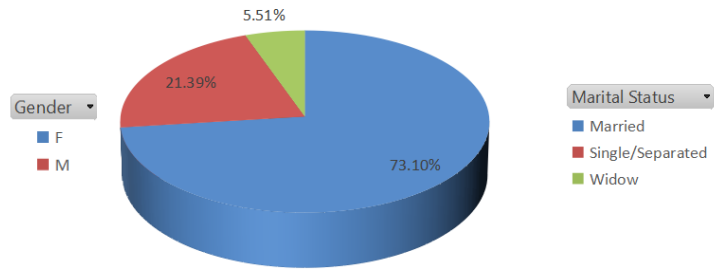


Figure 2

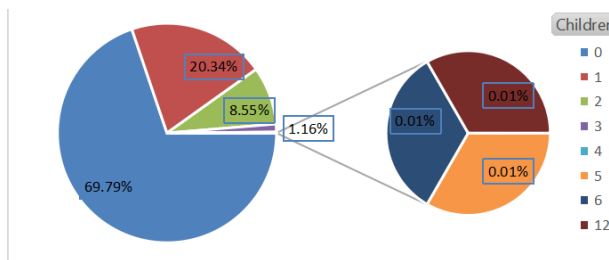


Figure 3

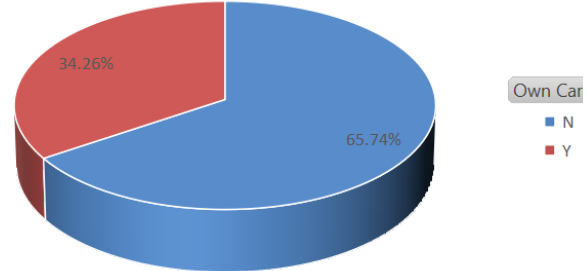


Figure 4

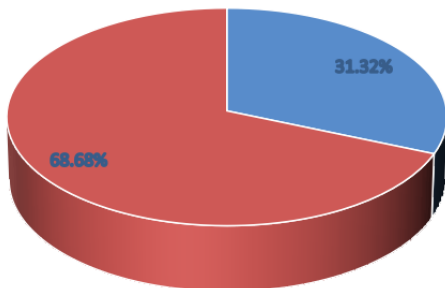


Figure 5

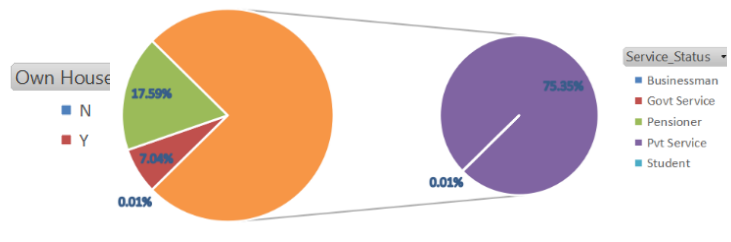


Figure 6

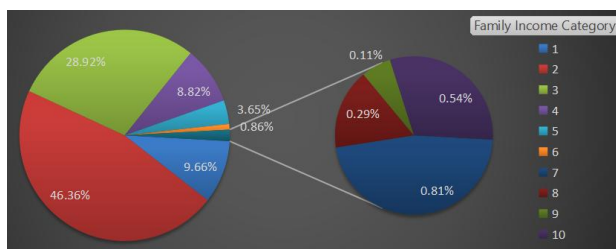


Figure 7

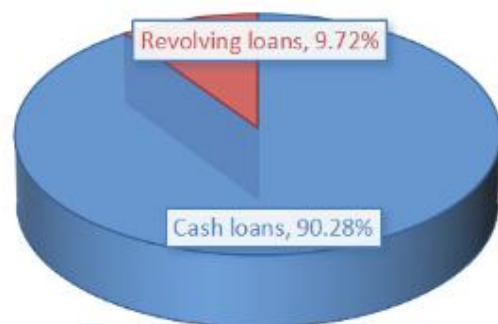


Figure 8

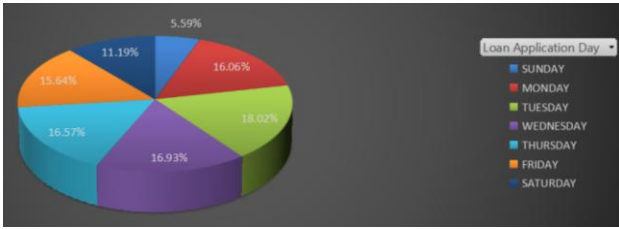


Figure 9

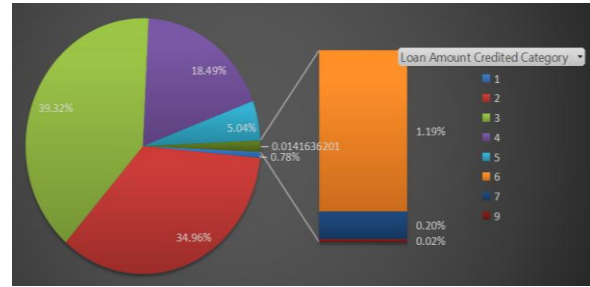


Figure 10

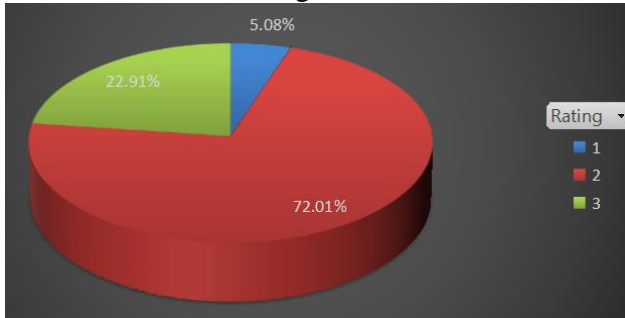


Figure 11

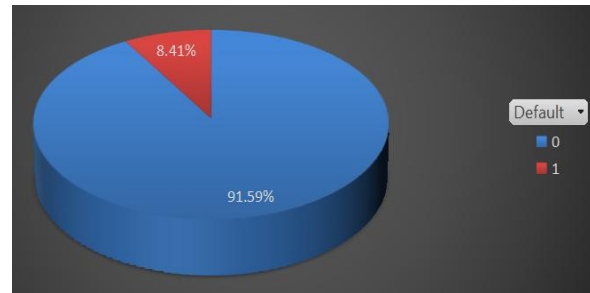


Figure 12

Table 1(a)
 Gender * Default Crosstabulation

			Default		Total
			0	1	
Gender	2	Count	9465	691	10156
		% within Gender	93.2%	6.8%	100.0%
		% within Default	67.0%	55.7%	66.1%
		% of Total	61.6%	4.5%	66.1%
1	Count	4669	550	5219	
	% within Gender	89.5%	10.5%	100.0%	
	% within Default	33.0%	44.3%	33.9%	
	% of Total	30.4%	3.6%	33.9%	
Total	Count	14134	1241	15375	
	% within Gender	91.9%	8.1%	100.0%	
	% within Default	100.0%	100.0%	100.0%	
	% of Total	91.9%	8.1%	100.0%	

Table 1(b)
 Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	-.065	.008	-8.066	.000 ^c
Ordinal by Ordinal	Spearman Correlation	-.065	.008	-8.066	.000 ^c
N of Valid Cases		15375			

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.
- c. Based on normal approximation.

Table 2(a)
Marital Status * Default Crosstabulation

			Default		Total
			0	1	
Marital Status	3	Count	808	39	847
		% within Marital Status	95.4%	4.6%	100.0%
		% within Default	5.7%	3.1%	5.5%
		% of Total	5.3%	.3%	5.5%
	2	Count	10344	895	11239
		% within Marital Status	92.0%	8.0%	100.0%
		% within Default	73.2%	72.1%	73.1%
		% of Total	67.3%	5.8%	73.1%
	1	Count	2982	307	3289
		% within Marital Status	90.7%	9.3%	100.0%
		% within Default	21.1%	24.7%	21.4%
		% of Total	19.4%	2.0%	21.4%
Total	Count	14134	1241	15375	
	% within Marital Status	91.9%	8.1%	100.0%	
	% within Default	100.0%	100.0%	100.0%	
	% of Total	91.9%	8.1%	100.0%	

Table 2(b)
Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	-.034	.008	-4.253	.000 ^c
Ordinal by Ordinal	Spearman Correlation	-.033	.008	-4.130	.000 ^c
N of Valid Cases		15375			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 3(a)
Children * Default Crosstabulation

			Default		Total
			0	1	
Children	12	Count	1	0	1
		% within Children	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%
	6	Count	1	0	1
		% within Children	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%

5	Count	1	0	1
	% within Children	100.0%	.0%	100.0%
	% within Default	.0%	.0%	.0%
	% of Total	.0%	.0%	.0%
4	Count	19	2	21
	% within Children	90.5%	9.5%	100.0%
	% within Default	.1%	.2%	.1%
	% of Total	.1%	.0%	.1%
3	Count	158	20	178
	% within Children	88.8%	11.2%	100.0%
	% within Default	1.1%	1.6%	1.2%
	% of Total	1.0%	.1%	1.2%
2	Count	1213	102	1315
	% within Children	92.2%	7.8%	100.0%
	% within Default	8.6%	8.2%	8.6%
	% of Total	7.9%	.7%	8.6%
1	Count	2846	282	3128
	% within Children	91.0%	9.0%	100.0%
	% within Default	20.1%	22.7%	20.3%
	% of Total	18.5%	1.8%	20.3%
0	Count	9895	835	10730
	% within Children	92.2%	7.8%	100.0%
	% within Default	70.0%	67.3%	69.8%
	% of Total	64.4%	5.4%	69.8%
Total	Count	14134	1241	15375
	% within Children	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 3(b)

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	.013	.008	1.552	.121 ^c
Ordinal by Ordinal Spearman Correlation	.015	.008	1.844	.065 ^c
N of Valid Cases	15375			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 4(a)
Own Car * Default Crosstabulation

			Default		Total
			0	1	
Own Car	1	Count	4858	409	5267
		% within Own Car	92.2%	7.8%	100.0%
		% within Default	34.4%	33.0%	34.3%
		% of Total	31.6%	2.7%	34.3%
0		Count	9276	832	10108
		% within Own Car	91.8%	8.2%	100.0%
		% within Default	65.6%	67.0%	65.7%
		% of Total	60.3%	5.4%	65.7%
Total		Count	14134	1241	15375
		% within Own Car	91.9%	8.1%	100.0%
		% within Default	100.0%	100.0%	100.0%
		% of Total	91.9%	8.1%	100.0%

Table 4(b)
Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	-.008	.008	-1.006	.314 ^c
Ordinal by Ordinal	Spearman Correlation	-.008	.008	-1.006	.314 ^c
N of Valid Cases		15375			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 5(a)
Own House * Default Crosstabulation

			Default		Total
			0	1	
Own House	1	Count	9728	831	10559
		% within Own House	92.1%	7.9%	100.0%
		% within Default	68.8%	67.0%	68.7%
		% of Total	63.3%	5.4%	68.7%
0		Count	4406	410	4816
		% within Own House	91.5%	8.5%	100.0%
		% within Default	31.2%	33.0%	31.3%
		% of Total	28.7%	2.7%	31.3%
Total		Count	14134	1241	15375
		% within Own House	91.9%	8.1%	100.0%
		% within Default	100.0%	100.0%	100.0%
		% of Total	91.9%	8.1%	100.0%

Table 5(b)
 Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	-.011	.008	-1.358	.174 ^c
Ordinal by Ordinal Spearman Correlation	-.011	.008	-1.358	.174 ^c
N of Valid Cases	15375			

- a. Not assuming the null hypothesis.
 b. Using the asymptotic standard error assuming the null hypothesis.
 c. Based on normal approximation.

Table 6(a)
 Service Status * Default Crosstabulation

		Default		Total	
		0	1		
Service Status	5	Count	1	0	1
		% within Service Status	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%
4	Count	1	0	1	
	% within Service Status	100.0%	.0%	100.0%	
	% within Default	.0%	.0%	.0%	
	% of Total	.0%	.0%	.0%	
3	Count	1014	69	1083	
	% within Service Status	93.6%	6.4%	100.0%	
	% within Default	7.2%	5.6%	7.0%	
	% of Total	6.6%	.4%	7.0%	
2	Count	2557	148	2705	
	% within Service Status	94.5%	5.5%	100.0%	
	% within Default	18.1%	11.9%	17.6%	
	% of Total	16.6%	1.0%	17.6%	
1	Count	10561	1024	11585	
	% within Service Status	91.2%	8.8%	100.0%	
	% within Default	74.7%	82.5%	75.3%	
	% of Total	68.7%	6.7%	75.3%	
Total	Count	14134	1241	15375	
	% within Service Status	91.9%	8.1%	100.0%	
	% within Default	100.0%	100.0%	100.0%	
	% of Total	91.9%	8.1%	100.0%	

Table 6(b)
 Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	-.043	.007	-5.330	.000 ^c
Ordinal by Ordinal Spearman Correlation	-.048	.007	-5.950	.000 ^c
N of Valid Cases	15375			

- a. Not assuming the null hypothesis.
 b. Using the asymptotic standard error assuming the null hypothesis.
 c. Based on normal approximation.

Table 7(a)
 FamilyIncomeCategory * Default Crosstabulation

			Default		Total
			0	1	
FamilyIncomeCategory	10	Count	18	0	18
		% within FamilyIncomeCategory	100.0%	.0%	100.0%
		% within Default	.1%	.0%	.1%
		% of Total	.1%	.0%	.1%
9	Count	4	0	4	
	% within FamilyIncomeCategory	100.0%	.0%	100.0%	
	% within Default	.0%	.0%	.0%	
	% of Total	.0%	.0%	.0%	
8	Count	12	0	12	
	% within FamilyIncomeCategory	100.0%	.0%	100.0%	
	% within Default	.1%	.0%	.1%	
	% of Total	.1%	.0%	.1%	
7	Count	36	3	39	
	% within FamilyIncomeCategory	92.3%	7.7%	100.0%	
	% within Default	.3%	.2%	.3%	
	% of Total	.2%	.0%	.3%	
6	Count	45	3	48	
	% within FamilyIncomeCategory	93.8%	6.2%	100.0%	
	% within Default	.3%	.2%	.3%	
	% of Total	.3%	.0%	.3%	
5	Count	230	15	245	
	% within FamilyIncomeCategory	93.9%	6.1%	100.0%	
	% within Default	1.6%	1.2%	1.6%	
	% of Total	1.5%	.1%	1.6%	
4	Count	690	51	741	

	% within FamilyIncomeCategory	93.1%	6.9%	100.0%
	% within Default	4.9%	4.1%	4.8%
	% of Total	4.5%	.3%	4.8%
3	Count	2953	285	3238
	% within FamilyIncomeCategory	91.2%	8.8%	100.0%
	% within Default	20.9%	23.0%	21.1%
	% of Total	19.2%	1.9%	21.1%
2	Count	7169	617	7786
	% within FamilyIncomeCategory	92.1%	7.9%	100.0%
	% within Default	50.7%	49.7%	50.6%
	% of Total	46.6%	4.0%	50.6%
1	Count	2977	267	3244
	% within Family Income Category	91.8%	8.2%	100.0%
	% within Default	21.1%	21.5%	21.1%
	% of Total	19.4%	1.7%	21.1%
Total	Count	14134	1241	15375
	% within FamilyIncomeCategory	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 7(b)
 Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	-.009	.007	-1.134	.257 ^c
Ordinal by Ordinal Spearman Correlation	-.002	.008	-.241	.810 ^c
N of Valid Cases	15375			

- a. Not assuming the null hypothesis.
 b. Using the asymptotic standard error assuming the null hypothesis.
 c. Based on normal approximation.

Table 8(a)
 Type Of Loan * Default Crosstabulation

		Default		Total	
		0	1		
Type Of Loan	2	Count	1423	72	1495
		% within Type Of Loan	95.2%	4.8%	100.0%
		% within Default	10.1%	5.8%	9.7%
		% of Total	9.3%	.5%	9.7%
	1	Count	12711	1169	13880

	% within Type Of Loan	91.6%	8.4%	100.0%
	% within Default	89.9%	94.2%	90.3%
	% of Total	82.7%	7.6%	90.3%
Total	Count	14134	1241	15375
	% within Type Of Loan	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 8(b)

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	-.039	.007	-4.867	.000 ^c
Ordinal by Ordinal Spearman Correlation	-.039	.007	-4.867	.000 ^c
N of Valid Cases	15375			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 9(a)

LoanApplicationDay * Default Crosstabulation

			Default		Total
			0	1	
Loan Application Day	7	Count	1570	150	1720
		% within Loan Application Day	91.3%	8.7%	100.0%
		% within Default	11.1%	12.1%	11.2%
		% of Total	10.2%	1.0%	11.2%
	6	Count	2199	206	2405
		% within Loan Application Day	91.4%	8.6%	100.0%
		% within Default	15.6%	16.6%	15.6%
		% of Total	14.3%	1.3%	15.6%
	5	Count	2357	191	2548
		% within Loan Application Day	92.5%	7.5%	100.0%
		% within Default	16.7%	15.4%	16.6%
		% of Total	15.3%	1.2%	16.6%
4	Count	2393	210	2603	
	% within Loan Application Day	91.9%	8.1%	100.0%	
	% within Default	16.9%	16.9%	16.9%	
	% of Total	15.6%	1.4%	16.9%	
3	Count	2539	232	2771	
	% within Loan Application Day	91.6%	8.4%	100.0%	

	% within Default	18.0%	18.7%	18.0%
	% of Total	16.5%	1.5%	18.0%
2	Count	2296	173	2469
	% within Loan Application Day	93.0%	7.0%	100.0%
	% within Default	16.2%	13.9%	16.1%
	% of Total	14.9%	1.1%	16.1%
1	Count	780	79	859
	% within Loan Application Day	90.8%	9.2%	100.0%
	% within Default	5.5%	6.4%	5.6%
	% of Total	5.1%	.5%	5.6%
Total	Count	14134	1241	15375
	% within Loan Application Day	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 9(b)

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	.008	.008	.972	.331 ^c
Ordinal by Ordinal Spearman Correlation	.008	.008	1.023	.306 ^c
N of Valid Cases	15375			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 10(a)

Loan Application Time * Default Crosstabulation

			Default		Total
			0	1	
Loan Application Time	23	Count	2	0	2
		% within Loan Application Time	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%
22	Count	7	2	9	
	% within Loan Application Time	77.8%	22.2%	100.0%	
	% within Default	.0%	.2%	.1%	
	% of Total	.0%	.0%	.1%	
21	Count	17	2	19	
	% within Loan Application Time	89.5%	10.5%	100.0%	
	% within Default	.1%	.2%	.1%	
	% of Total	.1%	.0%	.1%	

20	Count	49	5	54
	% within Loan Application Time	90.7%	9.3%	100.0%
	% within Default	.3%	.4%	.4%
	% of Total	.3%	.0%	.4%
19	Count	197	11	208
	% within Loan Application Time	94.7%	5.3%	100.0%
	% within Default	1.4%	.9%	1.4%
	% of Total	1.3%	.1%	1.4%
18	Count	428	32	460
	% within Loan Application Time	93.0%	7.0%	100.0%
	% within Default	3.0%	2.6%	3.0%
	% of Total	2.8%	.2%	3.0%
17	Count	653	54	707
	% within Loan Application Time	92.4%	7.6%	100.0%
	% within Default	4.6%	4.4%	4.6%
	% of Total	4.2%	.4%	4.6%
16	Count	991	74	1065
	% within Loan Application Time	93.1%	6.9%	100.0%
	% within Default	7.0%	6.0%	6.9%
	% of Total	6.4%	.5%	6.9%
15	Count	1158	76	1234
	% within Loan Application Time	93.8%	6.2%	100.0%
	% within Default	8.2%	6.1%	8.0%
	% of Total	7.5%	.5%	8.0%
14	Count	1222	99	1321
	% within Loan Application Time	92.5%	7.5%	100.0%
	% within Default	8.6%	8.0%	8.6%
	% of Total	7.9%	.6%	8.6%
13	Count	1394	135	1529
	% within Loan Application Time	91.2%	8.8%	100.0%
	% within Default	9.9%	10.9%	9.9%
	% of Total	9.1%	.9%	9.9%
12	Count	1578	165	1743
	% within Loan Application Time	90.5%	9.5%	100.0%
	% within Default	11.2%	13.3%	11.3%
	% of Total	10.3%	1.1%	11.3%
11	Count	1745	153	1898
	% within Loan Application Time	91.9%	8.1%	100.0%
	% within Default	12.3%	12.3%	12.3%
	% of Total	11.3%	1.0%	12.3%
10	Count	1729	153	1882
	% within Loan Application Time	91.9%	8.1%	100.0%

	% within Default	12.2%	12.3%	12.2%
	% of Total	11.2%	1.0%	12.2%
9	Count	1293	113	1406
	% within Loan Application Time	92.0%	8.0%	100.0%
	% within Default	9.1%	9.1%	9.1%
	% of Total	8.4%	.7%	9.1%
8	Count	658	64	722
	% within Loan Application Time	91.1%	8.9%	100.0%
	% within Default	4.7%	5.2%	4.7%
	% of Total	4.3%	.4%	4.7%
7	Count	412	37	449
	% within Loan Application Time	91.8%	8.2%	100.0%
	% within Default	2.9%	3.0%	2.9%
	% of Total	2.7%	.2%	2.9%
6	Count	267	23	290
	% within Loan Application Time	92.1%	7.9%	100.0%
	% within Default	1.9%	1.9%	1.9%
	% of Total	1.7%	.1%	1.9%
5	Count	163	23	186
	% within Loan Application Time	87.6%	12.4%	100.0%
	% within Default	1.2%	1.9%	1.2%
	% of Total	1.1%	.1%	1.2%
4	Count	95	11	106
	% within Loan Application Time	89.6%	10.4%	100.0%
	% within Default	.7%	.9%	.7%
	% of Total	.6%	.1%	.7%
3	Count	55	7	62
	% within Loan Application Time	88.7%	11.3%	100.0%
	% within Default	.4%	.6%	.4%
	% of Total	.4%	.0%	.4%
2	Count	16	1	17
	% within Loan Application Time	94.1%	5.9%	100.0%
	% within Default	.1%	.1%	.1%
	% of Total	.1%	.0%	.1%
1	Count	4	0	4
	% within Loan Application Time	100.0%	.0%	100.0%
	% within Default	.0%	.0%	.0%
	% of Total	.0%	.0%	.0%
0	Count	1	1	2
	% within Loan Application Time	50.0%	50.0%	100.0%
	% within Default	.0%	.1%	.0%
	% of Total	.0%	.0%	.0%

Total	Count	14134	1241	15375
	% within Loan Application Time	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 10(b)
 Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	-.023	.008	-2.806	.005 ^c
Ordinal by Ordinal Spearman Correlation	-.021	.008	-2.613	.009 ^c
N of Valid Cases	15375			

- a. Not assuming the null hypothesis.
 b. Using the asymptotic standard error assuming the null hypothesis.
 c. Based on normal approximation.

Table 11(a)
 Loan Amount Credited * Default Crosstabulation

			Default		Total
			0	1	
Loan Amount Credited	9	Count	1	0	1
		% within Loan Amount Credited	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%
7	Count	11	1	12	
	% within Loan Amount Credited	91.7%	8.3%	100.0%	
	% within Default	.1%	.1%	.1%	
	% of Total	.1%	.0%	.1%	
6	Count	78	4	82	
	% within Loan Amount Credited	95.1%	4.9%	100.0%	
	% within Default	.6%	.3%	.5%	
	% of Total	.5%	.0%	.5%	
5	Count	396	20	416	
	% within Loan Amount Credited	95.2%	4.8%	100.0%	
	% within Default	2.8%	1.6%	2.7%	
	% of Total	2.6%	.1%	2.7%	
4	Count	1780	129	1909	
	% within Loan Amount Credited	93.2%	6.8%	100.0%	
	% within Default	12.6%	10.4%	12.4%	
	% of Total	11.6%	.8%	12.4%	
3	Count	4933	481	5414	
	% within Loan Amount Credited	91.1%	8.9%	100.0%	
	% within Default	34.9%	38.8%	35.2%	
	% of Total	32.1%	3.1%	35.2%	

2	Count	6628	591	7219
	% within Loan Amount Credited	91.8%	8.2%	100.0%
	% within Default	46.9%	47.6%	47.0%
	% of Total	43.1%	3.8%	47.0%
1	Count	307	15	322
	% within Loan Amount Credited	95.3%	4.7%	100.0%
	% within Default	2.2%	1.2%	2.1%
	% of Total	2.0%	.1%	2.1%
Total	Count	14134	1241	15375
	% within Loan Amount Credited	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 11(b)
 Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	-.013	.007	-1.599	.110 ^c
Ordinal by Ordinal Spearman Correlation	-.007	.008	-.814	.416 ^c
N of Valid Cases	15375			

- a. Not assuming the null hypothesis.
 b. Using the asymptotic standard error assuming the null hypothesis.
 c. Based on normal approximation.

Table 12(a)
 Actual Price Loan Item * Default Crosstabulation

			Default		Total
			0	1	
Actual Price Loan Item	9	Count	1	0	1
		% within Actual Price Loan Item	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%
	7	Count	1	0	1
		% within Actual Price Loan Item	100.0%	.0%	100.0%
		% within Default	.0%	.0%	.0%
		% of Total	.0%	.0%	.0%
	6	Count	37	1	38
		% within Actual Price Loan Item	97.4%	2.6%	100.0%
		% within Default	.3%	.1%	.2%
		% of Total	.2%	.0%	.2%
	5	Count	272	16	288
		% within Actual Price Loan Item	94.4%	5.6%	100.0%

	% within Default	1.9%	1.3%	1.9%
	% of Total	1.8%	.1%	1.9%
4	Count	1256	59	1315
	% within Actual Price Loan Item	95.5%	4.5%	100.0%
	% within Default	8.9%	4.8%	8.6%
	% of Total	8.2%	.4%	8.6%
3	Count	4236	354	4590
	% within Actual Price Loan Item	92.3%	7.7%	100.0%
	% within Default	30.0%	28.5%	29.9%
	% of Total	27.6%	2.3%	29.9%
2	Count	7885	783	8668
	% within Actual Price Loan Item	91.0%	9.0%	100.0%
	% within Default	55.8%	63.1%	56.4%
	% of Total	51.3%	5.1%	56.4%
1	Count	432	28	460
	% within Actual Price Loan Item	93.9%	6.1%	100.0%
	% within Default	3.1%	2.3%	3.0%
	% of Total	2.8%	.2%	3.0%
Total	Count	14120	1241	15361
	% within Actual Price Loan Item	91.9%	8.1%	100.0%
	% within Default	100.0%	100.0%	100.0%
	% of Total	91.9%	8.1%	100.0%

Table 12(b)
 Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	-.040	.007	-4.983	.000 ^c
Ordinal by Ordinal	Spearman Correlation	-.038	.007	-4.661	.000 ^c
N of Valid Cases		15361			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 13(a)
 Region Rating * Default Crosstabulation

			Default		Total
			0	1	
Region Rating	3	Count	2158	252	2410
		% within Region Rating	89.5%	10.5%	100.0%
		% within Default	15.3%	20.3%	15.7%
		% of Total	14.0%	1.6%	15.7%
	2	Count	10453	910	11363
		% within Region Rating	92.0%	8.0%	100.0%
		% within Default	74.0%	73.3%	73.9%
		% of Total	68.0%	5.9%	73.9%
	1	Count	1523	79	1602
		% within Region Rating	95.1%	4.9%	100.0%
		% within Default	10.8%	6.4%	10.4%
		% of Total	9.9%	.5%	10.4%
Total	Count	14134	1241	15375	
	% within Region Rating	91.9%	8.1%	100.0%	
	% within Default	100.0%	100.0%	100.0%	
	% of Total	91.9%	8.1%	100.0%	

Table 13(b)
 Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	.051	.008	6.288	.000 ^c
Ordinal by Ordinal	Spearman Correlation	.051	.008	6.270	.000 ^c
N of Valid Cases		15375			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Table 14
 ANOVA

Default with Marital Status	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.556	2	.778	10.495	.000
Within Groups	1139.277	15372	.074		
Total	1140.832	15374			

Table 15
ANOVA

Default with Service Status					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.837	4	.709	9.581	.000
Within Groups	1137.995	15370	.074		
Total	1140.832	15374			

Table 16
ANOVA

Default with Family Income					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.634	9	.070	.949	.481
Within Groups	1140.199	15365	.074		
Total	1140.832	15374			

Table 17
ANOVA

Default with Loan Amount Credited					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.605	7	.229	3.094	.003
Within Groups	1139.227	15367	.074		
Total	1140.832	15374			

Table 18
ANOVA

Default with Region Rating					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.955	2	1.477	19.960	.000
Within Groups	1137.877	15372	.074		
Total	1140.832	15374			

Table 19
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.584
Bartlett's Test of Sphericity	Approx. Chi-Square
	df
	Sig.
	3.483E4
	78
	.000

Table 20
 Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.286	17.586	17.586	2.286	17.586	17.586	2.147	16.515	16.515
2	1.515	11.657	29.243	1.515	11.657	29.243	1.470	11.306	27.821
3	1.316	10.119	39.363	1.316	10.119	39.363	1.361	10.467	38.288
4	1.100	8.464	47.827	1.100	8.464	47.827	1.112	8.551	46.839
5	1.035	7.963	55.790	1.035	7.963	55.790	1.101	8.468	55.307
6	1.017	7.821	63.611	1.017	7.821	63.611	1.080	8.304	63.611
7	.984	7.567	71.177						
8	.873	6.715	77.893						
9	.814	6.260	84.152						
10	.720	5.537	89.689						
11	.624	4.802	94.491						
12	.615	4.728	99.219						
13	.102	.781	100.000						

Extraction Method: Principal Component Analysis.

Table 21
 Component Matrix^a

	Component					
	1	2	3	4	5	6
Loan Amount Credited	.885					
Actual Price Loan Item	.878					
FamilyIncomeCategory	.633					
Gender		.655				
OwnCar		-.578				
ServiceStatus						
Loan Application Time			-.716			
Region Rating			.663			
OwnHouse				.728		
Type Of Loan				.507		
Children					.686	
MaritalStatus					.584	
Loan Application Day						.615

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

Table 22
 Rotated Component Matrix^a

	Component					
	1	2	3	4	5	6
Loan Amount Credited	.938					
Actual Price Loan Item	.932					
FamilyIncomeCategory	.532					
Gender		.796				
OwnCar		-.750				
Loan Application Time			.784			
Region Rating			-.758			
OwnHouse				.785		
Type Of Loan				.561		
Children					.797	
ServiceStatus					-.510	
Loan Application Day						
MaritalStatus						.844

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Table 23
 Component Transformation Matrix

Component	1	2	3	4	5	6
1	.921	-.283	.237	-.120	.027	.005
2	.310	.793	-.260	-.080	-.370	.251
3	.117	-.350	-.892	.002	.129	.228
4	.106	.055	.097	.924	.153	.316
5	-.067	.156	.185	-.344	.649	.630
6	.161	.375	-.194	.088	.633	-.622

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

Table 24
 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.584
Bartlett's Test of Sphericity	Approx. Chi-Square
	df
	Sig.
	3.480E4
	66
	.000

Table 25
 Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.286	19.049	19.049	2.286	19.049	19.049	2.139	17.827	17.827
2	1.515	12.622	31.671	1.515	12.622	31.671	1.447	12.057	29.885
3	1.315	10.959	42.630	1.315	10.959	42.630	1.353	11.275	41.159
4	1.097	9.139	51.769	1.097	9.139	51.769	1.109	9.245	50.404
5	1.035	8.626	60.395	1.035	8.626	60.395	1.107	9.222	59.627
6	1.004	8.364	68.759	1.004	8.364	68.759	1.096	9.132	68.759
7	.873	7.275	76.034						
8	.814	6.784	82.818						
9	.720	6.003	88.822						
10	.625	5.210	94.032						
11	.615	5.122	99.153						
12	.102	.847	100.000						

Extraction Method: Principal Component Analysis.

Table 26
 Component Matrix^a

	Component					
	1	2	3	4	5	6
Loan Amount Credited	.885					
Actual Price Loan Item	.878					
FamilyIncomeCategory	.633					
Gender		.655				
OwnCar		-.580				
Loan Application Time			-.715			
Region Rating			.665			
OwnHouse				.740		
Type Of Loan				.505		
Children					.692	
MaritalStatus					.571	
ServiceStatus						-.532

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

Table 27
 Rotated Component Matrix^a

	Component					
	1	2	3	4	5	6
Loan Amount Credited	.944					
Actual Price Loan Item	.939					
FamilyIncomeCategory	.534					
Gender		-.820				
OwnCar		.779				
Loan Application Time			.786			
Region Rating			-.776			
OwnHouse				.807		
Type Of Loan				.598		
Children					.855	
ServiceStatus					-.558	
MaritalStatus						.878

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Table 28
 Component Transformation Matrix

Component	1	2	3	4	5	6
1	.920	.283	.237	-.126	.029	.000
2	.295	-.765	-.243	-.112	-.396	.315
3	.111	.359	-.888	-.022	.125	.233
4	.097	-.010	.104	.900	.092	.401
5	-.060	-.175	.177	-.335	.719	.552
6	.201	-.418	-.234	.220	.549	-.617

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

Table 29
 Correlations

	REGR factor score 1 for analysis 1	REGR factor score 2 for analysis 1	REGR factor score 3 for analysis 1	REGR factor score 4 for analysis 1	REGR factor score 5 for analysis 1	REGR factor score 6 for analysis 1
REGR factor score 1 for analysis 1	1	.000	.000	.000	.000	.000
Pearson Correlation		1.000	1.000	1.000	1.000	1.000
Sig. (2-tailed)						
N	15361	15361	15361	15361	15361	15361
REGR factor score 2 for analysis 1	.000	1	.000	.000	.000	.000
Pearson Correlation						

analysis 1	Sig. (2-tailed)	1.000		1.000	1.000	1.000	1.000
	N	15361	15361	15361	15361	15361	15361
REGR factor score 3 for analysis 1	Pearson Correlation	.000	.000	1	.000	.000	.000
	Sig. (2-tailed)	1.000	1.000		1.000	1.000	1.000
	N	15361	15361	15361	15361	15361	15361
REGR factor score 4 for analysis 1	Pearson Correlation	.000	.000	.000	1	.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000		1.000	1.000
	N	15361	15361	15361	15361	15361	15361
REGR factor score 5 for analysis 1	Pearson Correlation	.000	.000	.000	.000	1	.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000		1.000
	N	15361	15361	15361	15361	15361	15361
REGR factor score 6 for analysis 1	Pearson Correlation	.000	.000	.000	.000	.000	1
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	1.000	
	N	15361	15361	15361	15361	15361	15361

Table 30
 Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
REGR factor score 1 for analysis 1	15361	-2.48983	7.96637	.0000000	1.0000000
REGR factor score 2 for analysis 1	15361	-3.39824	1.92002	.0000000	1.0000000
REGR factor score 3 for analysis 1	15361	-3.21718	3.99674	.0000000	1.0000000
REGR factor score 4 for analysis 1	15361	-2.49277	3.78308	.0000000	1.0000000
REGR factor score 5 for analysis 1	15361	-2.86623	12.02779	.0000000	1.0000000
REGR factor score 6 for analysis 1	15361	-2.65625	4.61354	0.000000	1.0000000
Valid N (listwise)	15361				

Block 0: Beginning Block

Table 31
 Classification Table^{a,b}

Observed			Predicted		
			Default		Percentage Correct
			0	1	
Step 0	Default	0	14120	0	100.0
		1	1241	0	.0
Overall Percentage					91.9

a. Constant is included in the model.
 b. The cut value is .500

Table 32
 Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-2.432	.030	6.745E3	1	.000	.088

Table 33
 Variables not in the Equation

	Score	df	Sig.
Step 0 Variables			
Gender(1)	64.913	1	.000
Marital Status	21.050	2	.000
Marital Status(1)	9.101	1	.003
Marital Status(2)	.688	1	.407
Children	2.433	1	.119
Own Car(1)	1.061	1	.303
Own House(1)	1.839	1	.175
Service Status	38.337	4	.000
Service Status(1)	37.399	1	.000
Service Status(2)	30.000	1	.000
Service Status(3)	4.540	1	.033
Service Status(4)	.088	1	.767
Family Income Category	8.532	9	.482
Family Income Category(1)	.145	1	.704
Family Income Category(2)	.460	1	.497
Family Income Category(3)	2.928	1	.087
Family Income Category(4)	1.475	1	.225
Family Income Category(5)	1.283	1	.257
Family Income Category(6)	.217	1	.641
Family Income Category(7)	.008	1	.929
Family Income Category(8)	1.055	1	.304
Family Income Category(9)	.352	1	.553
Type Of Loan(1)	22.846	1	.000
Loan Application Day	8.447	6	.207
Loan Application Day(1)	1.559	1	.212
Loan Application Day(2)	4.500	1	.034
Loan Application Day(3)	.408	1	.523
Loan Application Day(4)	.000	1	.987
Loan Application Day(5)	1.324	1	.250
Loan Application Day(6)	.935	1	.334
Loan Application Time	7.774	1	.005
Loan Amount Credited	2.643	1	.104
Actual Price Loan Item	24.790	1	.000
Region Rating	39.385	1	.000

Variables not in the Equation

Step 0	Variables	Score	df	Sig.
	Gender(1)	64.913	1	.000
	Marital Status	21.050	2	.000
	Marital Status(1)	9.101	1	.003
	Marital Status(2)	.688	1	.407
	Children	2.433	1	.119
	Own Car(1)	1.061	1	.303
	Own House(1)	1.839	1	.175
	Service Status	38.337	4	.000
	Service Status(1)	37.399	1	.000
	Service Status(2)	30.000	1	.000
	Service Status(3)	4.540	1	.033
	Service Status(4)	.088	1	.767
	Family Income Category	8.532	9	.482
	Family Income Category(1)	.145	1	.704
	Family Income Category(2)	.460	1	.497
	Family Income Category(3)	2.928	1	.087
	Family Income Category(4)	1.475	1	.225
	Family Income Category(5)	1.283	1	.257
	Family Income Category(6)	.217	1	.641
	Family Income Category(7)	.008	1	.929
	Family Income Category(8)	1.055	1	.304
	Family Income Category(9)	.352	1	.553
	Type Of Loan(1)	22.846	1	.000
	Loan Application Day	8.447	6	.207
	Loan Application Day(1)	1.559	1	.212
	Loan Application Day(2)	4.500	1	.034
	Loan Application Day(3)	.408	1	.523
	Loan Application Day(4)	.000	1	.987
	Loan Application Day(5)	1.324	1	.250
	Loan Application Day(6)	.935	1	.334
	Loan Application Time	7.774	1	.005
	Loan Amount Credited	2.643	1	.104
	Actual Price Loan Item	24.790	1	.000
	Region Rating	39.385	1	.000
	Overall Statistics	254.449	30	.000

Block 1: Method = Enter

Table 34

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	259.636	30	.000
	Block	259.636	30	.000
	Model	259.636	30	.000

Table 35
Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.773	8	.673

Table 36
Contingency Table for Hosmer and Lemeshow Test

		Default = .00		Default = 1.00		Total
		Observed	Expected	Observed	Expected	
Step 1	1	1485	1.486E3	51	49.996	1536
	2	1460	1.465E3	76	70.502	1536
	3	1446	1.453E3	90	83.333	1536
	4	1453	1.440E3	83	95.876	1536
	5	1428	1.427E3	108	108.586	1536
	6	1417	1.414E3	119	121.634	1536
	7	1395	1.399E3	140	135.738	1535
	8	1386	1.383E3	150	153.378	1536
	9	1370	1.356E3	166	180.311	1536
	10	1280	1.296E3	258	241.645	1538

Table 37
Classification Table^a

Observed			Predicted		Percentage Correct
			Default		
			0	1	
Step 1	Default	0	14120	0	100.0
		1	1241	0	.0
Overall Percentage					91.9

a. The cut value is .500

Table 38
 Variables in the Equation

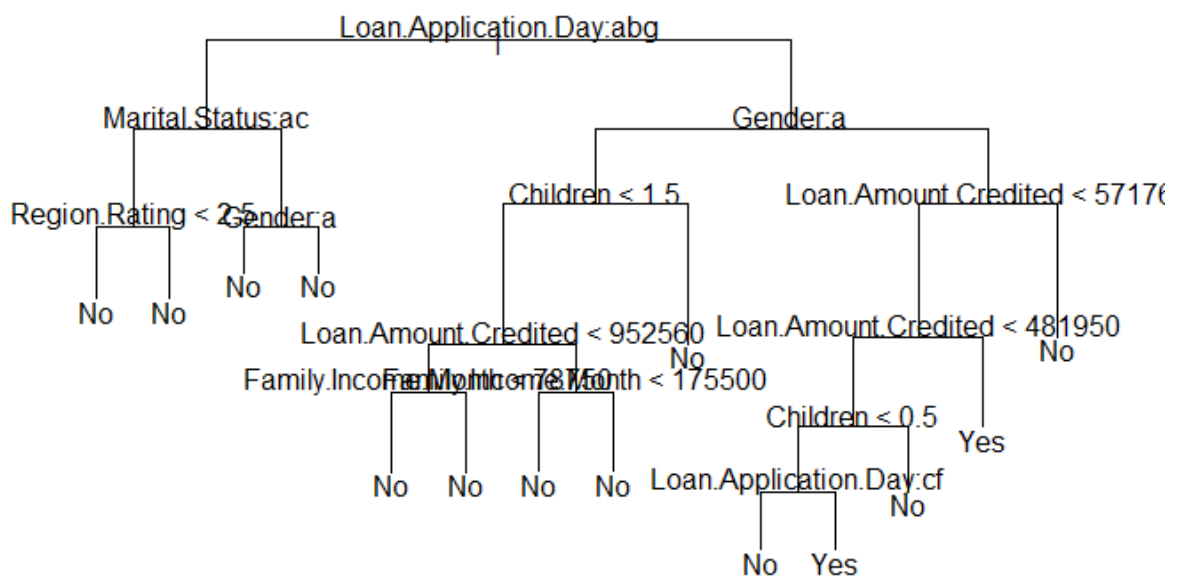
	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a								
Gender(1)	.497	.066	56.376	1	.000	1.644	1.444	1.872
Marital Status			10.610	2	.005			
Marital Status(1)	.509	.180	7.974	1	.005	1.663	1.168	2.367
Marital Status(2)	.339	.174	3.791	1	.052	1.404	.998	1.974
Children	.022	.043	.262	1	.609	1.022	.940	1.111
Own Car(1)	.265	.069	14.598	1	.000	1.303	1.138	1.493
Own House(1)	.023	.064	.126	1	.723	1.023	.902	1.161
Service Status			23.817	4	.000			
Service Status(1)	17.525	4.001E4	.000	1	1.000	4.084E7	.000	.
Service Status(2)	17.075	4.001E4	.000	1	1.000	2.604E7	.000	.
Service Status(3)	17.266	4.001E4	.000	1	1.000	3.152E7	.000	.
Service Status(4)	-.394	5.672E4	.000	1	1.000	.674	.000	.
Family Income Category			4.596	9	.868			
Family Income Category(1)	18.562	9.232E3	.000	1	.998	1.152E8	.000	.
Family Income Category(2)	18.463	9.232E3	.000	1	.998	1.043E8	.000	.
Family Income Category(3)	18.604	9.232E3	.000	1	.998	1.201E8	.000	.
Family Income Category(4)	18.442	9.232E3	.000	1	.998	1.022E8	.000	.
Family Income Category(5)	18.431	9.232E3	.000	1	.998	1.010E8	.000	.
Family Income Category(6)	18.478	9.232E3	.000	1	.998	1.059E8	.000	.
Family Income Category(7)	18.785	9.232E3	.000	1	.998	1.439E8	.000	.
Family Income Category(8)	.023	1.458E4	.000	1	1.000	1.024	.000	.
Family Income Category(9)	.029	2.192E4	.000	1	1.000	1.030	.000	.
Type Of Loan(1)	.574	.128	20.189	1	.000	1.776	1.382	2.281
Loan Application Day			8.082	6	.232			
Loan Application Day(1)	.067	.147	.209	1	.648	1.070	.801	1.428
Loan Application Day(2)	-.239	.118	4.141	1	.042	.787	.625	.991
Loan Application Day(3)	-.043	.111	.154	1	.695	.958	.771	1.190
Loan Application Day(4)	-.071	.113	.393	1	.531	.932	.747	1.163
Loan Application Day(5)	-.123	.115	1.148	1	.284	.884	.705	1.108
Loan Application Day(6)	.002	.114	.000	1	.987	1.002	.802	1.251
Loan Application Time	-.008	.010	.732	1	.392	.992	.973	1.011

Loan Amount Credited	.374	.071	27.437	1	.000	1.453	1.264	1.671
Actual Price Loan Item	-.550	.080	47.721	1	.000	.577	.493	.674
Region Rating	.343	.062	30.544	1	.000	1.410	1.248	1.592
Constant	-39.856	4.105E4	.000	1	.999	.000		

a. Variable(s) entered on step 1: Gender, Marital Status, Children, Own Car, Own House, Service Status, Family Income Category, Type Of Loan, Loan Application Day, Loan Application Time, Loan Amount Credited, Actual Price Loan Item, Region Rating.

Machine Learning Output

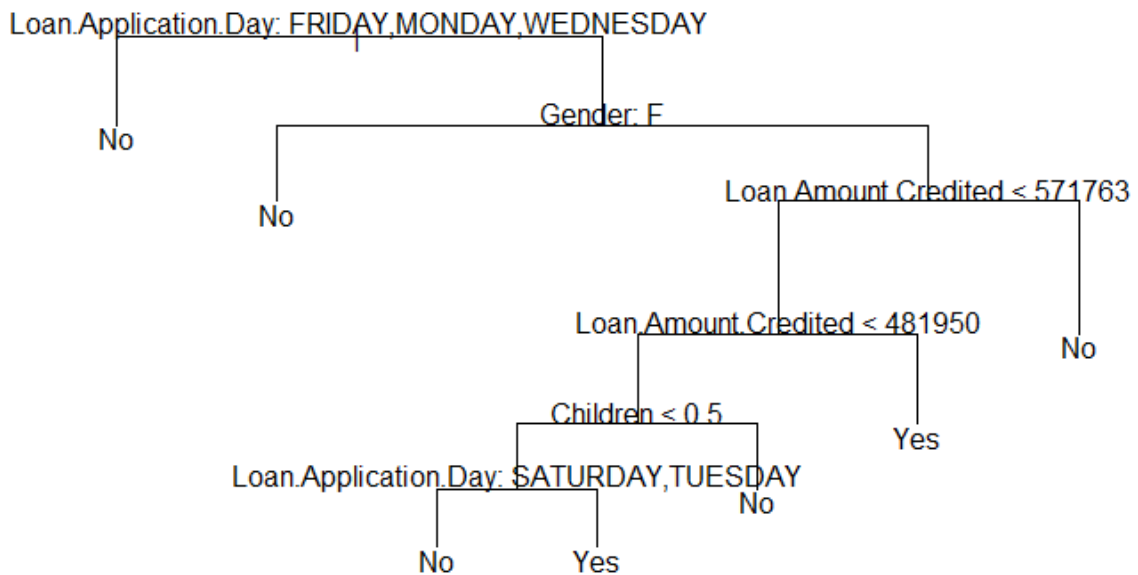
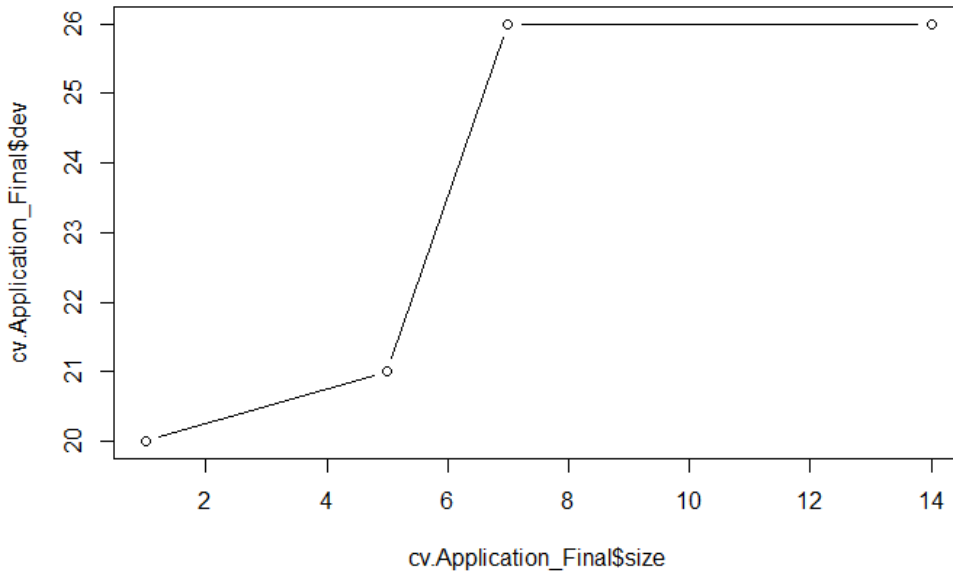
Initial Tree



```

> table(tree.pred,Default.test)
Default.test
tree.pred   No   Yes
No  13167  1122
Yes    772   100
    
```

Figure 13



```
> table(tree.pred,Default.test)
```

Default.test		No	Yes
tree.pred	No	13410	1142
tree.pred	Yes	529	80

Bagging

```
> bag.Application_Final
```

Call:

```
randomForest(formula = Default ~ Gender + Marital.Status + Own.Car +
Own.House + Service.Status + Family.Income.Month + Type.of.Loan +
Loan.Application.Day + Loan.Application.Time + Loan.Amount.Credited +
Actual.Price.of.Loan.Item + Region.Rating, mtry = 14, importance = TRUE,
subset = train)
```

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 12

OOB estimate of error rate: 8.33%

Confusion matrix:

```
No      Yes  class.error
No  9167  25  0.002719756
Yes   808   0  1.000000000
```

```
># Prediction
>yhat.bag = predict(bag.Application_Final,newdata=Application_Final.test)
># Confusion Matrix
>table(yhat.bag,Default.test)
Default.test
yhat.bagNo  Yes
No  4912  433
Yes  16   0
Random Forest
>rf.Application_Final
Call:
randomForest(formula = Default ~ Gender + Marital.Status + Children +
Own.Car + Own.House + Service_Status + Family.Income.Month +
Type.of.Loan + Loan.Application.Day + Loan.Application.Time +
Loan.Amount.Credited + Actual.Price.of.Loan.Item + Region.Rating,
data = Application_Final, mtry = 4, importance = TRUE, subset = train)
Type of random forest: classification
Number of trees: 500
No. of variables tried at each split: 4
OOB estimate of error rate: 8.12%
```

```
Confusion matrix:
No  Yes  class.error
No  9188  4  0.000435161
Yes  808   0  1.000000000
```

```
># Prediction
>yhat.rf = predict(rf.Application_Final,newdata=Application_Final.test)
># Confusion Matrix
>table(yhat.rf,Default.test)
Default.test
yhat.rf  No  Yes
No  4923  433
Yes   5   0
># Importance of the Variables
>importance(rf.Application_Final,type="2")
MeanDecreaseGini
Gender                18.190472
Marital.Status       42.359294
Children             67.007050
Own.Car              36.138154
Own.House            41.432680
Service_Status       26.029913
Family.Income.Month 243.507052
Type.of.Loan         8.955837
Loan.Application.Day 153.697608
Loan.Application.Time 200.701094
Loan.Amount.Credited 300.798388
Actual.Price.of.Loan.Item 220.508487
Region.Rating        48.024859
```

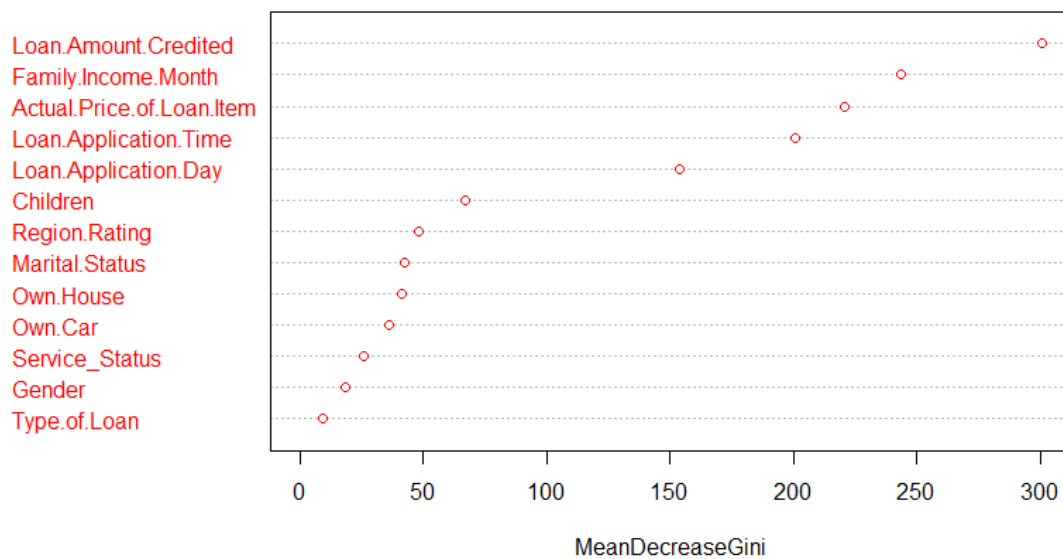


Figure 14

Table 39 (Comparative Performance)

Technique	Error Rate
Logistic Regression	8%
Tree	11.2%
Bagging	8.33%
Random Forest	8.1%