MATH 271.1 Credit Scorecard Project

Brian Godwin S. Lim,
Terence Brian O. Tsai

1 Introduction

Credit scorecards are one of the popular models used to evaluate the credit risk associated with applicants based on the information about the applicant. In its essence, a scorecard consists of a group of characteristics about the applicant which are statistically proven to be predictive in separating good and bad accounts. Applicants are given scores based on these characteristics and are generally allowed to take out loans only if they exceed a certain score. Credit scorecards are preferred by many since they are easy to interpret, explain, implement, and use compared to other blackbox models.

Using data on 23,337 individual loan accounts, we developed two credit scorecard models, segmented based on the purpose of the loan. Information about the days past due (DPD) of each account were given as well as 13 characteristics of the client of each account. Before performing any analysis or computation, the dataset was first cleaned and outliers were removed. We also checked for any invalid entries and dealt with them accordingly. Afterwards, the development of the credit scorecard started with defining the bad definition and confirming it using roll-rate analysis and current versus worst delinquency comparison method. This was used to assign the creditworthiness of each account in the dataset.

Following this, we segmented the dataset based on the purpose of the loan. An exploratory data analysis was then conducted to check the correlation of creditworthiness with each variable. Afterwards, we split the dataset into training and test datasets. Using the training dataset, we determined the logical binnings of each variable based on the weights of evidence (WOE) as well as the information value (IV). The results are then validated against the test dataset to determine if the logical binnings still holds. After confirming the logical binnings, we replaced all raw data values in the training and test datasets with their corresponding WOEs and dropped the unpredictive variables, as well as sex to avoid gender discrimination issues.

Using the training dataset, we performed logistic regression and created different models using forward, backward, and step-wise regression for our variable selection process. The different models were then compared with each other. We then determined the optimal cutoff threshold using two different objective functions. Using these thresholds, we evaluated the different models against the test dataset to determine the best model based on the sensitivity. Several other metrics were also taken into consideration. Finally, we converted the model into a scorecard with the appropriate cutoff score.

2 Dataset

2.1 Overview of the Dataset

The dataset consists of 23,337 individual loan accounts. The days past due (DPD) of each account account for the historical window October 1, 2019 to October 1, 2020 were given. This indicates the number of months that the account was delinquent in paying the amount due for a given month. Furthermore, we are also given 13 information about each applicant.

1. `Account`: account number
2. `Sex`: biological sex
3. `Dependents`: number of dependents
4. `Civil_Status`: civil status
5. `House_Type`: house type
6. `Education`: highest educational attainment
7. `Yrs_Employed`: number of years employed (from first to current employment)
8. `Credit_Status`: status whether the credit account has current (or ongoing) loan, non-earning (or non-paying) loan, or paid-off loan
9. `Months_Loan`: duration of the loan in months
10. `Amortization`: amortization amount per month
11. `Purpose_Loan`: purpose/reason for loaning
12. `Gross_Salary`: average gross salary per month
13. `Credit_Ratio`: calculated by dividing the amortization by the gross salary

The objective of the credit scorecard model is to use these characteristics to create a model that could predict the creditworthiness of the accounts in the dataset, as well as future applicants.

2.2 Cleaning the Dataset

Figure 1 in the following page shows the boxplot of Dependents, Months Loan, and Years Employed. It can be seen that there are outliers in these variables. Thus, we removed these accounts from the dataset to avoid any problems in our analysis moving forward. Specifically, we removed accounts with more than 10 dependents, more than 49 years employed, or more than 72 months loan. After removing these outliers, we are left with 23,321 accounts.

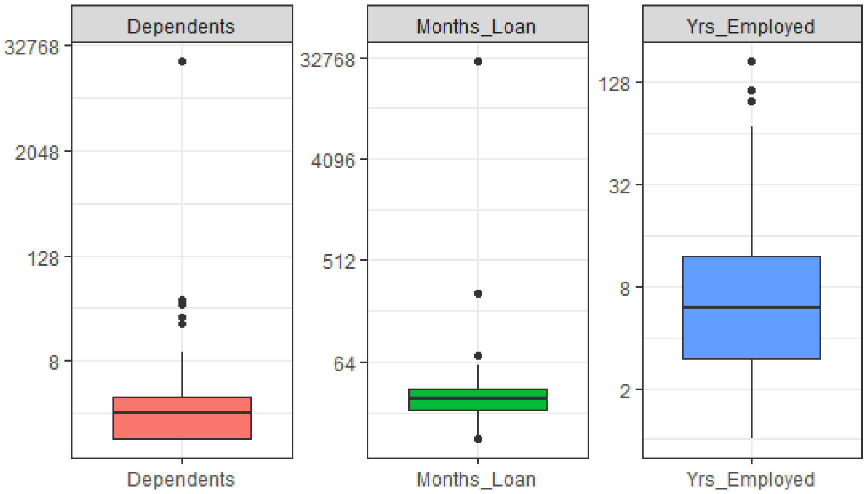

Figure 1: Boxplot of Dependents, Months Loan, Years Employed

Additionally, using the `DataExplorer::plot_missing()` function, we checked for invalid entries in the dataset. We can see that there were no invalid (`NA`) entries in the dataset. However, we do note that under the House Type category, some of the data were encoded as "missing info." In our analysis, missing info was considered as one of the categories for this variable.

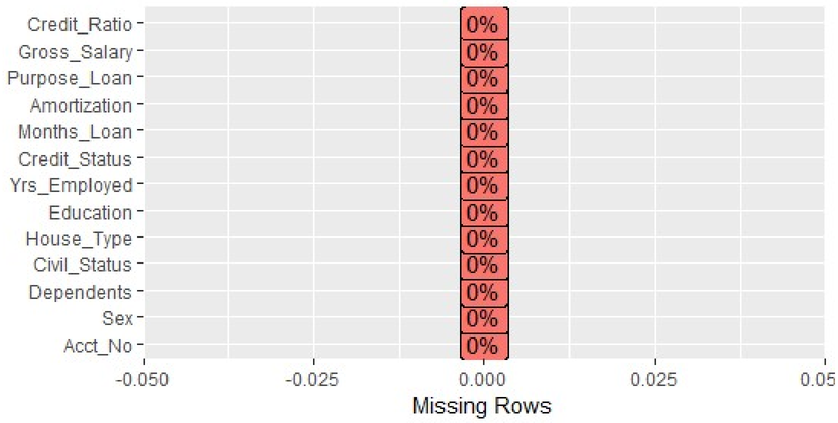

Figure 2: Missing Values

3 Bad Definition

In order to classify each account as a good or bad account, we needed an objective definition of a "bad" account. In this paper, 90 days past due (DPD) was used as this definition, consistent with the definition given in the Basel II Capital Accord. We confirmed this "bad" definition of 90 days past due (DPD) by performing a roll-rate analysis as well as current versus worst delinquency comparison on the historical delinquency performance of the accounts.

3.1 Roll-Rate Analysis

In roll-rate analysis, we want to compare the worst delinquency in the first x months with that in the next x months. Afterwards, we calculate the percentage of accounts that maintain their worst delinquency, get better, or roll forward into the next delinquency brackets. The objective is to determine the number of days past due (DPD) after which the account is highly likely to default as the client is unable to revert back to regular payments for the loan.

Using the dataset, we compared the worst delinquency in the first 6 months with that in the next 6 months. Figure 3 below shows the contingency table where the rows are the worst delinquency in the first 6 months and the columns are the worst delinquency in the next 6 months. The table shows that there are 6,506 roll forward accounts representing 27.90% of all accounts.

	roll_rate_2												
roll_rate_1	0	1	2	3	4	5	6	7	8	9	10	11	12
0	15302	682	305	160	148	117	127	0	0	0	0	0	0
1	196	95	85	43	27	31	28	109	0	0	0	0	0
2	94	27	22	67	48	23	15	10	134	0	0	0	2
3	434	25	111	453	383	108	188	104	84	2461	0	0	0
4	21	2	3	4	4	12	6	0	4	7	153	0	0
5	1	0	1	0	0	1	4	0	6	3	3	149	0
6	3	0	0	0	1	1	0	8	7	1	5	166	0
7	3	0	0	0	0	0	0	3	3	5	0	7	122
8	4	0	0	0	0	2	0	0	0	4	0	2	142
9	0	0	0	0	0	0	2	0	0	0	0	0	88
10	0	0	0	0	0	0	0	0	0	0	0	0	59
11	0	0	0	0	0	0	0	0	0	0	0	0	51

Figure 3: Contingency Table

	roll_rate_2												
roll_rate_1	0	1	2	3	4	5	6	7	8	9	10	11	12
0	90.86	4.05	1.81	0.95	0.88	0.69	0.75	0.00	0.00	0.00	0.00	0.00	0.00
1	31.92	15.47	13.84	7.00	4.40	5.05	4.56	17.75	0.00	0.00	0.00	0.00	0.00
2	21.27	6.11	4.98	15.16	10.86	5.20	3.39	2.26	30.32	0.00	0.00	0.00	0.45
3	9.97	0.57	2.55	10.41	8.80	2.48	4.32	2.39	1.93	56.56	0.00	0.00	0.00
4	9.72	0.93	1.39	1.85	1.85	5.56	2.78	0.00	1.85	3.24	70.83	0.00	0.00
5	0.60	0.00	0.60	0.00	0.00	0.60	2.38	0.00	3.57	1.79	1.79	88.69	0.00
6	1.56	0.00	0.00	0.00	0.52	0.52	0.00	4.17	3.65	0.52	2.60	86.46	0.00
7	2.10	0.00	0.00	0.00	0.00	0.00	0.00	2.10	2.10	3.50	0.00	4.90	85.31
8	2.60	0.00	0.00	0.00	0.00	1.30	0.00	0.00	0.00	2.60	0.00	1.30	92.21
9	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.00	0.00	0.00	0.00	0.00	97.78
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Figure 4: Transition Matrix

Converting the contingency table into percentages, we obtain the transition matrix shown above. Using this matrix, we can obtain the probability of maintaining or worsening given the worst delinquency in the first 6 months below. We observe that the first level of delinquency in the first 6 months such that more than 75% of accounts maintained or worsened their delinquency is three months pass due or 90 DPD. This means that once the client is 90 DPD, there is a 86.90% chance that they will maintain or worsen in their delinquency. Thus, the roll-rate analysis confirms the bad definition of 90 DPD.

0	1	2	3	4	5	6	7	8	9	10	11
100.00	68.08	72.62	86.90	86.11	98.81	97.40	97.90	96.10	97.78	100.00	100.00

Figure 5: Roll Forward Probabilities

3.2 Current versus Worst Delinquency Comparison

The current versus worst delinquency comparison, on the other hand, focuses on the comparison of the worst ever delinquency status of accounts in the past with their most current delinquency status. Similar to the roll rate analysis, the objective is also to look for a “point of no return” where we classify an account as bad.

Using the dataset, we compared the worst ever delinquency in the previous 11 months with that in the current (12^{th}) month. Figure 6 below shows the worst versus current table. The rows indicate the worst delinquency in the previous 11 months while the columns indicate the delinquency in the current (12^{th}) month. Similarly, we can convert the table to obtain a worst versus current matrix of percentages as shown in Figure 7 below.

worst \ current	0	1	2	3	4	5	6	7	8	9	10	11	12
0	15302	210	0	0	0	0	0	0	0	0	0	0	0
1	656	107	194	0	0	0	0	0	0	0	0	0	0
2	246	45	48	151	0	0	0	0	0	0	0	0	0
3	545	74	94	429	550	0	0	0	0	0	0	0	0
4	38	4	30	8	10	214	0	0	0	0	0	0	0
5	57	2	2	1	7	11	277	0	0	0	0	0	0
6	63	0	28	2	0	0	3	228	0	0	0	0	0
7	3	0	3	3	0	0	0	0	228	0	0	0	0
8	12	1	0	0	0	0	0	0	3	2472	0	0	0
9	7	2	0	0	1	0	0	0	1	0	159	0	0
10	1	0	0	0	0	0	0	0	0	1	0	155	0
11	0	0	0	0	0	0	0	0	0	0	0	169	122
12	1	0	0	0	0	0	0	0	0	0	0	0	341

Figure 6: Worst versus Current Table

worst \ current	0	1	2	3	4	5	6	7	8	9	10	11	12
0	98.65	1.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	68.55	11.18	20.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	50.20	9.18	9.80	30.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	32.21	4.37	5.56	25.35	32.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	12.50	1.32	9.87	2.63	3.29	70.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	15.97	0.56	0.56	0.28	1.96	3.08	77.59	0.00	0.00	0.00	0.00	0.00	0.00
6	19.44	0.00	8.64	0.62	0.00	0.00	0.93	70.37	0.00	0.00	0.00	0.00	0.00
7	1.27	0.00	1.27	1.27	0.00	0.00	0.00	0.00	96.20	0.00	0.00	0.00	0.00
8	0.48	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.12	99.36	0.00	0.00	0.00
9	4.12	1.18	0.00	0.00	0.59	0.00	0.00	0.00	0.59	0.00	93.53	0.00	0.00
10	0.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.64	0.00	98.73	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	58.08	41.92
12	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	99.71

Figure 7: Worst versus Current Matrix

From the matrix above, we obtain the percent of accounts with a given worst delinquency for the previous 11 months worsened. Figure 8 below shows the said percentages. We see that 57.86% of all accounts that has a 90-day delinquency stayed at 90 days or became worse. Since the 90 DPD is the first delinquency bucket to reach the threshold of 50%, the current versus worst comparison also confirms that the 90 DPD bad definition.

0	1	2	3	4	5	6	7	8	9	10	11	12
100.00	31.45	40.61	57.86	73.68	80.67	71.30	96.20	99.48	93.53	98.73	100.00	99.71

Figure 8: Maintain or Worsen Percentages

Therefore, we can classify accounts having 90 days past due or more as bad accounts while accounts having less than 90 days past due as good accounts. Doing so, we get that 27.93% of all accounts are bad.

4 Exploratory Data Analysis

After confirming the bad definition, we perform exploratory data analysis on the entire dataset. We compared the distribution of each variable across good and bad accounts as well as the overall dataset to gain insights on the correlation of the variable and creditworthiness.

4.1 Account Number

Figure 9 shows that the good and bad accounts appear to be equally distributed across the account numbers. The mean of bad accounts is 11580 with a variance of 44588992. The mean of good accounts is 11703 with a variance of 45691174. The mean of all accounts is 11668 with a variance of 45384475. These numbers confirm that the good and bad accounts are equally distributed among the range of account numbers. Thus, we can infer that account number does not have a strong correlation with creditworthiness. This is as expected as account numbers are generally assigned in chronological order to those who open accounts, and as such, we expect no great disparity in terms of default rate among the ranges of accounts.

Figure 9: EDA: Account Number

4.2 Sex

In Figure 10 below, it is evident that there are approximately equal number of both sexes for bad accounts. However, for good accounts, there were significantly more male than female. Putting this to perspective with the overall data where it is observed that there are significantly more male than female, this seems to indicate that female accounts are more likely to default than male accounts Thus, sex might have a strong correlation with creditworthiness. However, we must also be cautious in including sex to our credit scorecard model as this might give rise to issues of gender discrimination. As such, we chose to omit this variable in creating our model.

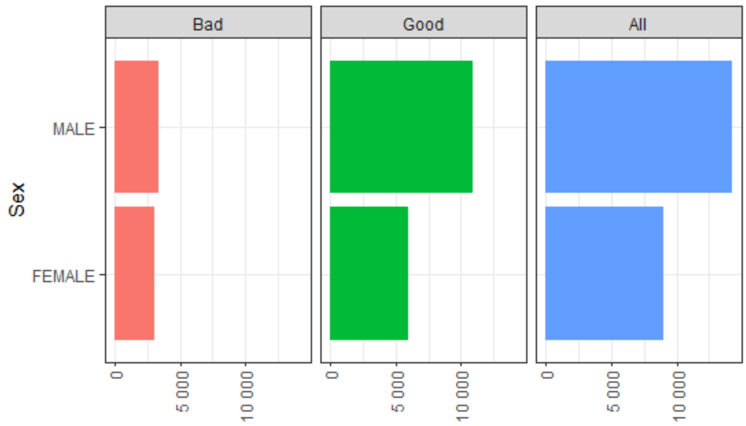

Figure 10: EDA: Sex

4.3 Dependents

In terms of the number of dependents, Figure 11 points to good accounts generally having more dependents. Bad accounts have a mean of 1.31 dependents with a variance of 1.46 while good accounts have a mean of 1.36 dependents with a variance of 2.24. Across all accounts, we get a mean of 1.34 dependents with a variance of 2.02. These values further indicates that good accounts have slightly more dependents than bad accounts. Thus, we cannot discount that the number of dependents might have a high correlation with creditworthiness.

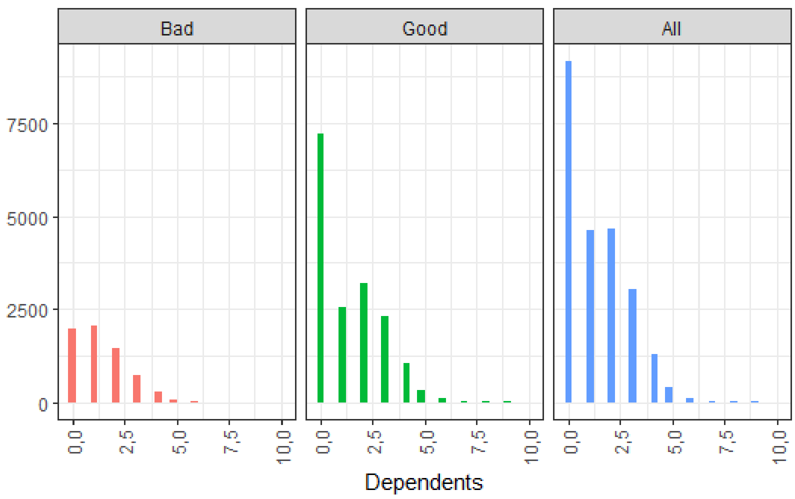

Figure 11: EDA: Dependents

4.4 Civil Status

Looking at civil status in Figure 12, all four statuses have approximately the same distribution in both good and bad accounts. In both cases, married was the largest followed by single in terms of the number of accounts. Both widowed and separated have only few accounts. Thus, it appears that civil status has a weak correlation with creditworthiness.

Figure 12: EDA: Civil Status

4.5 House Type

In Figure 13, we looked at house type. We see a clear distinction in owned houses. There were significantly fewer bad accounts with owned houses. This is as expected since we can expect those who own their homes to be more financially stable, and as such, have better capability to pay back their loans. For the other house types, the proportion of defaults did not differ much in both good and bad accounts. Thus, house types, especially the difference between owning your own house or not, can potentially have a high correlation with creditworthiness.

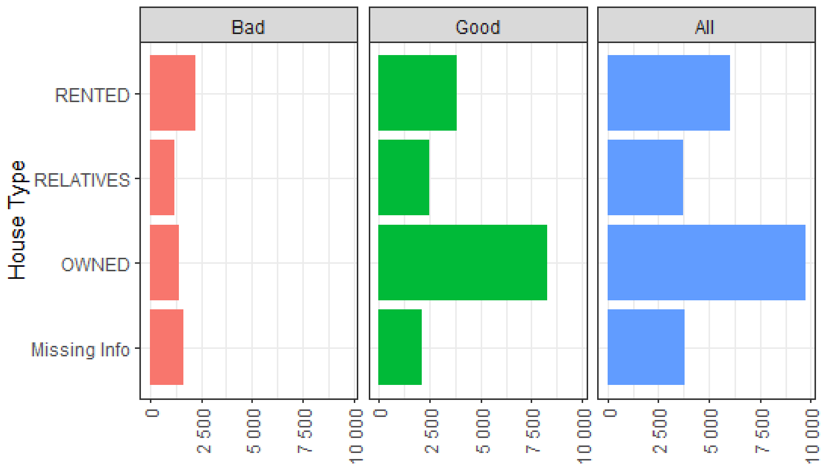

Figure 13: EDA: House Type

4.6 Education

In terms of education, Figure 14 highlights that there were only a few good accounts whose highest education was elementary. Moreover, majority of accounts whose highest education was college was classified as good accounts. The large discrepancy points towards education having a strong correlation with creditworthiness.

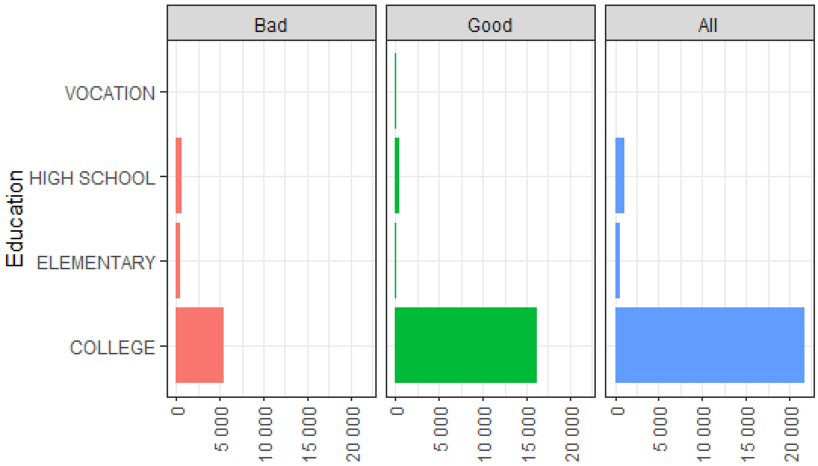

Figure 14: EDA: Education

4.7 Years Employed

Taking a close look at the number of years employed in Figure 15, we can see that for good accounts, there were more accounts with more years employed. However, bad accounts have a greater mean years employed. Bad accounts have a mean of 8.72 years and variance of 29.2 while good accounts have a mean of 7.92 years and variance of 44.3. Overall the mean is 8.15 years with variance of 40.2. We must take into account that good accounts have a greater variance. Moreover, even as good accounts have a lower mean, logic dictates that higher employment years means more stability and this fact, along with the observation that among those at the tail end of the number of years are have a greater percentage of good accounts tell us that number of years could possibly be positively correlated with creditworthiness.

Figure 15: EDA: Years Employed

4.8 Credit Status

In Figure 16, credit status was analyzed. We immediately see that most of the paid-off accounts are good while most of the non-earning accounts are bad. For the current accounts, we see more good accounts than bad accounts as one might expect. Thus, it can be inferred that credit status has a strong correlation with creditworthiness.

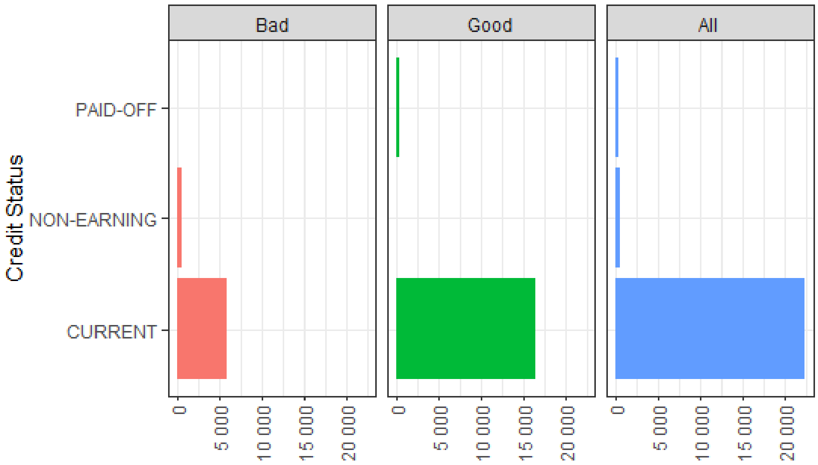

Figure 16: EDA: Credit Status

4.9 Months Loan

Similarly, in Figure 17, it appears that bad accounts have a larger number of months loan compared to good accounts. Looking at the mean, bad accounts have a mean of 31.1 months with a variance of 138 while good accounts have a mean of 30.0 months with a variance of 93.8. Overall, we have a mean of 30.6 months and a variance of 107. This is as expected as more risk is associated with loans that have longer duration. Thus, the months loan seems to have a large correlation with creditworthiness.

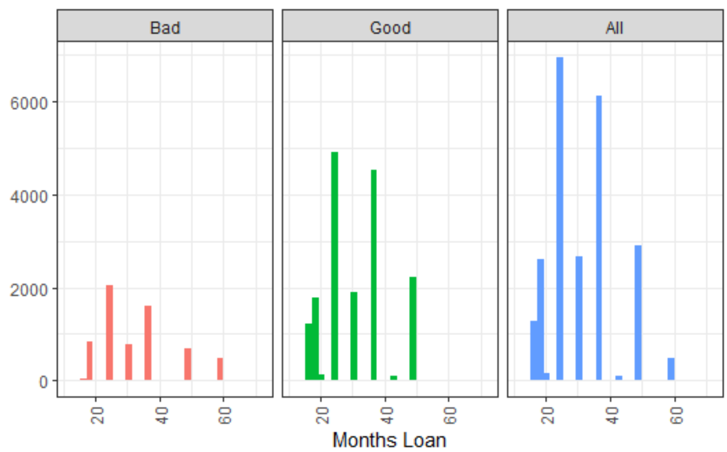

Figure 17: EDA: Months Loan

4.10 Amortization

Inspecting amortization in Figure 18, good accounts reaches higher amortization amounts compared to bad accounts. Bad accounts have a mean of 4621 and a variance of 28255556 while good accounts have a mean of 4670 and a variance of 45734455. Across all accounts, we have a mean of 4656 and a variance of 40852079. We must note that good accounts have a significantly higher variance compared to bad accounts. Therefore, we cannot discount that amortization might have a high correlation with creditworthiness.

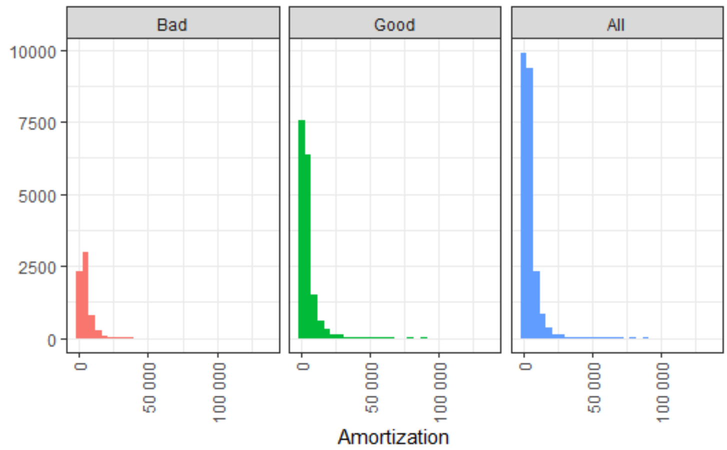

Figure 18: EDA: Amortization

4.11 Purpose of Loan

A look at the purpose of the loan in Figure 19 reveals that for medicinal purpose, there were more bad accounts than good accounts. While for the other purposes, there were more good accounts. Therefore, we need to consider that the purpose of the loan could possibly affect creditworthiness.

Figure 19: EDA: Purpose of Loan

4.12 Gross Salary

Looking at gross salary in Figure 20, we see that good accounts reaches higher gross salary compared to bad accounts. Bad accounts have a mean of 27545 with a variance of 533777139 while good accounts have a mean of 31845 and a variance of 1481074668. Across all accounts, we have a mean of 30644 and a variance of 1220203665. Again, we must point out that good accounts have a significantly mean and higher variance. The higher mean of good accounts is as expected as those who get paid more are expected to be able to pay back their loans more punctually and fully. Thus, gross salary can have a large correlation with creditworthiness.

Figure 20: EDA: Gross Salary

4.13 Credit Ratio

Looking at credit ratio in Figure 21 below, we see that bad accounts reaches higher credit ratio compared to good accounts. Bad accounts have a mean of 0.162 with a variance of 0.00639 while good accounts have a mean of 0.147 and a variance of 0.00346. The higher. mean of bad accounts is expected as those with higher utilization of credit lines are expected to have worse creditworthiness than those with lower utilization. Overall, we have a mean of 0.151 and a variance of 0.00433. Similarly, we must point out that bad accounts have about double the variance of good accounts. Hence, we need to inspect how credit ratio affects creditworthiness.

Figure 21: EDA: Credit Ratio

5 The Segmented Credit Scorecard - Essentials

For the first credit scorecard model, we only consider accounts where the purpose of loan is essential. We only include accounts where the purpose of loan is either home repair, tuition, medical, or to pay debts. After segmentation, we are left with 17,400 accounts 5,412 of which are bad accounts representing 31.10% of the essentials accounts.

5.1 Exploratory Data Analysis

Next, we perform exploratory data analysis on the entire segmented dataset. We compared the distribution of each variable across good and bad accounts as well as the overall population to gain insights on the correlation of the variable and creditworthiness.

5.1.1 Account Number

First, we take a look at Account number. Figure 22 shows that the good and bad accounts appear to be equally distributed across the account numbers and there are no observable trends. The mean of bad accounts is 11669 with a variance of 44457151. The mean of good accounts is 11695 with a variance of 46180689. The mean of all accounts is 11687 with a variance of 45642172. These numbers confirm that the good and bad accounts are equally distributed among the range of account numbers. Thus, we can infer that account number does not have a strong correlation with creditworthiness. This is as expected as, as stated above, account numbers are generally assigned in chronological order to those who open accounts, and as such, we expect no great disparity in terms of default rate among the ranges of accounts.

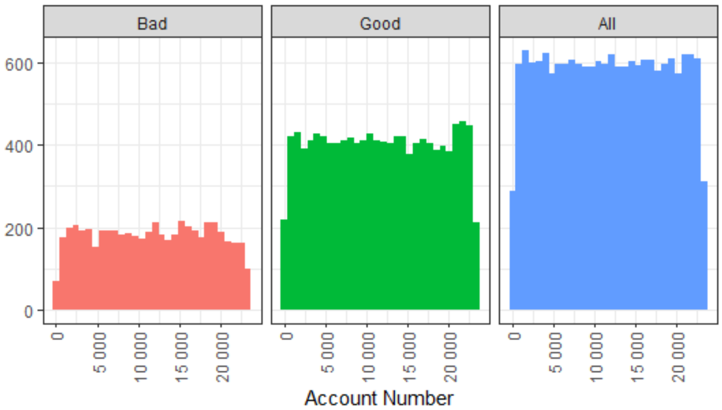

Figure 22: EDA: Account Number

5.1.2 Sex

In Figure 23 below, it is evident that there are approximately equal number of both sexes for bad accounts. However, for good accounts, there were significantly more male than female. Putting this to perspective with the overall data where it is observed that there are significantly more male than female, this seems to indicate that female accounts are more likely to default than male accounts Thus, sex might have a strong correlation with creditworthiness. However, we must also be cautious in including sex to our credit scorecard model as this might give rise to issues of gender discrimination. As such, we chose to omit this variable in creating our model.

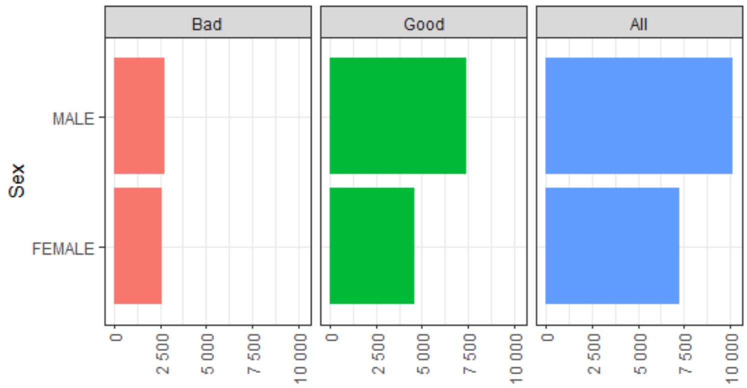

Figure 23: EDA: Sex

5.1.3 Dependents

In terms of the number of dependents, Figure 24 points to good accounts generally having more dependents. Bad accounts have a mean of 1.31 dependents with a variance of 1.45 while good accounts have a mean of 1.41 dependents with a variance of 2.26. Across all accounts, we get a mean of 1.38 dependents with a variance of 2.01. These values further indicates that good accounts have slightly more dependents than bad accounts. This can be attributed to how financially stable families can afford to have more dependents. Thus, we cannot discount that the number of dependents might have a high correlation with creditworthiness.

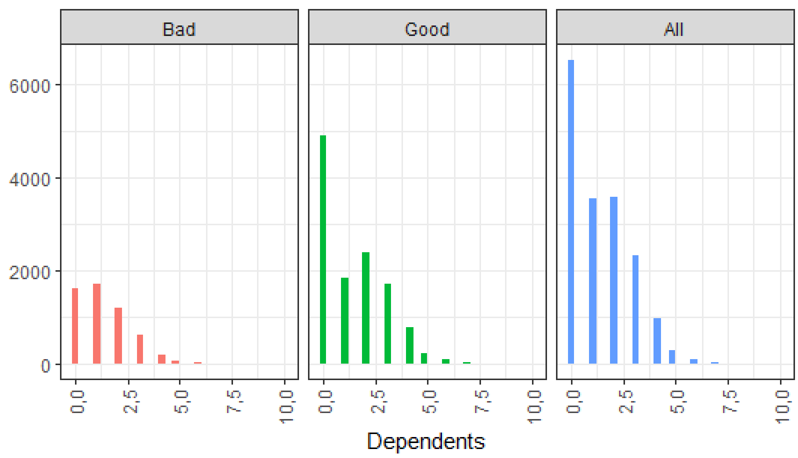

Figure 24: EDA: Dependents

5.1.4 Civil Status

Looking at civil status in Figure 25, all four statuses have approximately the same distribution in both good and bad accounts. In both cases, married was the largest followed by single in terms of the number of accounts. Both widowed and separated have only few accounts. This can be explained by how on average, married clients have more sources of income to shoulder financial responsibilities and single clients have less financial responsibilities. Thus, it appears that civil status has a weak correlation with creditworthiness.

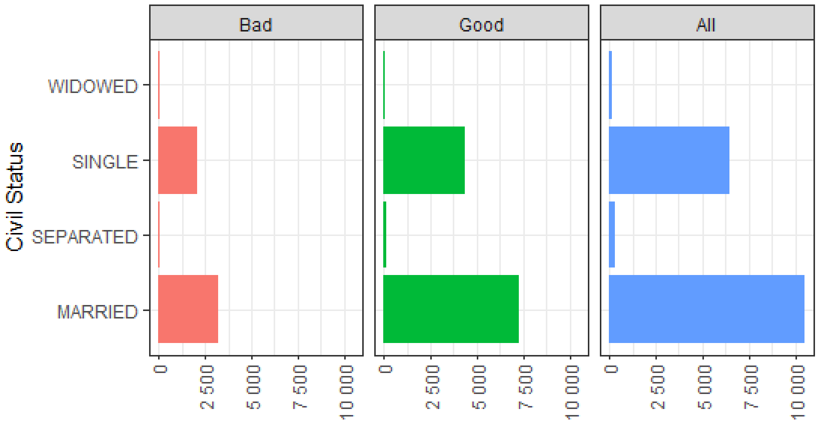

Figure 25: EDA: Civil Status

5.1.5 House Type

In Figure 26, we looked at house type. We see a clear distinction in owned houses. There were significantly fewer bad accounts with owned houses. This is as expected since we can expect those who own their homes to be more financially stable, and as such, have better capability to pay back their loans. For the other house types, the proportion of defaults did not differ much in both good and bad accounts. Thus, house types, especially the difference between owning your own house or not, can potentially have a high correlation with creditworthiness.

Figure 26: EDA: House Type

5.1.6 Education

In terms of education, Figure 27 highlights that there were only a few good accounts whose highest education was elementary. Moreover, majority of accounts whose highest education was college was classified as good accounts. This trend appears to be consistent with how college and vocation graduates generally have better job opportunities that allow them to pay back their loans. The large discrepancy points towards education having a strong correlation with creditworthiness.

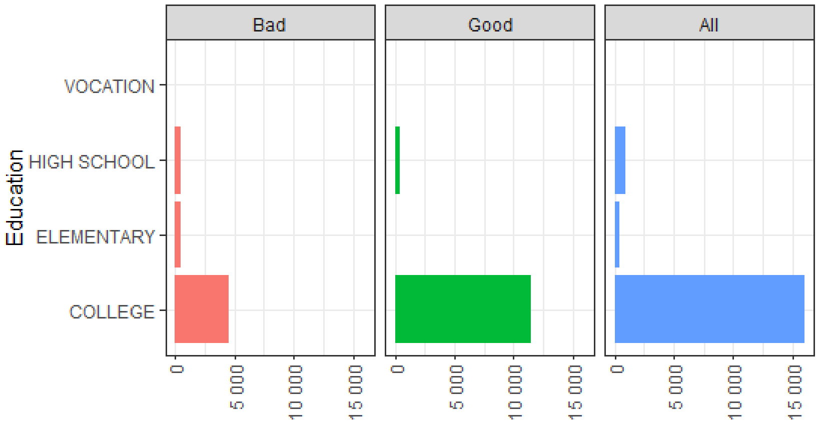

Figure 27: EDA: Education

5.1.7 Years Employed

Taking a close look at the number of years employed in Figure 28, we can see that for good accounts, there were more accounts with more years employed. However, bad accounts have a greater mean years employed. Bad accounts have a mean of 8.77 years and variance of 28.5 while good accounts have a mean of 8.16 years and variance of 45.2. Overall the mean is 8.35 years with variance of 40.0. We must take into account that good accounts have a greater variance. Thus, there appears to be a non-linear relationship between the number of years employed and creditworthiness. Therefore, we cannot discount the correlation between these variables.

Figure 28: EDA: Years Employed

5.1.8 Credit Status

In Figure 29, credit status was analyzed. We immediately see that most of the paid-off accounts are good while most of the non-earning accounts are bad. For the current accounts, we see more good accounts than bad accounts as one might expect. Thus, it can be inferred that credit status has a strong correlation with creditworthiness.

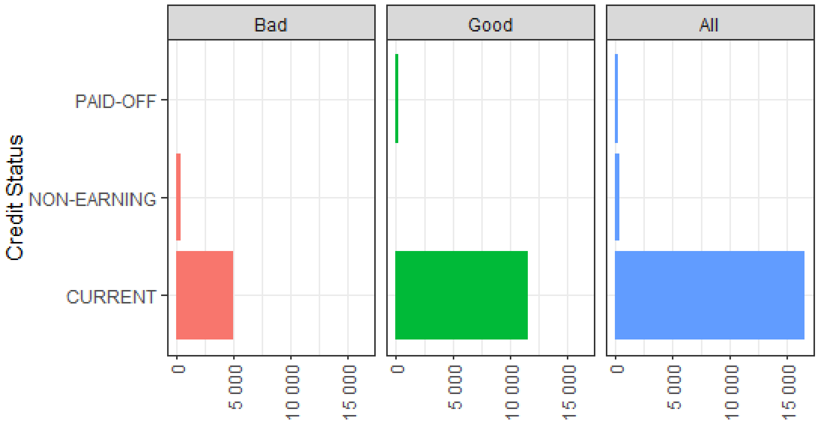

Figure 29: EDA: Credit Status

5.1.9 Months Loan

Similarly, in Figure 30, it appears that bad accounts have a larger number of months loan compared to good accounts. Looking at the mean, bad accounts have a mean of 32.4 months with a variance of 142 while good accounts have a mean of 30.9 months with a variance of 86.9. Overall, we have a mean of 31.3 months and a variance of 105. This is as expected as more risk is associated with loans that have longer duration. Thus, the months loan seems to have a large correlation with creditworthiness.

Figure 30: EDA: Months Loan

5.1.10 Amortization

Inspecting amortization in Figure 31, good accounts reaches higher amortization amounts compared to bad accounts. Bad accounts have a mean of 4398 and a variance of 15634932 while good accounts have a mean of 4108 and a variance of 33957039. Across all accounts, we have a mean of 4198 and a variance of 28274999. We must note that good accounts have a significantly higher variance compared to bad accounts. This is in line with our expectations since higher monthly amortization means larger financial obligations. Therefore, we cannot discount that amortization might have a high correlation with creditworthiness.

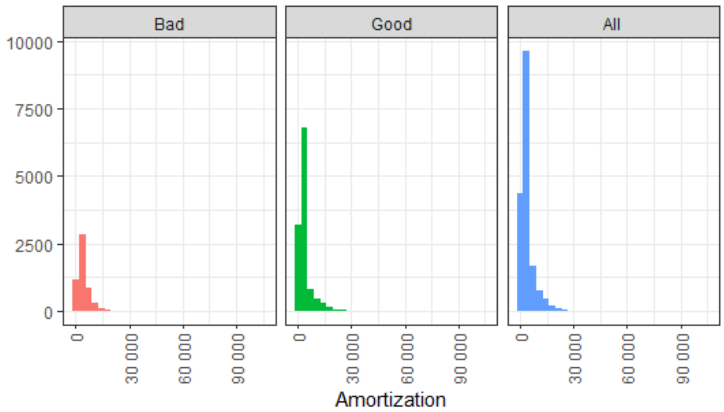

Figure 31: EDA: Amortization

5.1.11 Purpose of Loan

A look at the purpose of the loan in Figure 32 reveals that for medicinal purpose, there were more bad accounts than good accounts. This is as expected loans for medicinal purposes are usually come unexpected and clients would find this harder to pay back. While for the other purposes, there were more good accounts. Therefore, we need to consider that the purpose of the loan could possibly affect creditworthiness.

Figure 32: EDA: Purpose of Loan

5.1.12 Gross Salary

Looking at gross salary in Figure 33, we see that good accounts reaches higher gross salary compared to bad accounts. Bad accounts have a mean of 26586 with a variance of 329124411 while good accounts have a mean of 27885 and a variance of 1087923373. Across all accounts, we have a mean of 27481 and a variance of 852239872. Again, we must point out that good accounts have a higher mean and higher variance. The higher mean of good accounts is as expected as those who get paid more are expected to be able to pay back their loans more punctually and fully. Thus, gross salary can have a large correlation with creditworthiness.

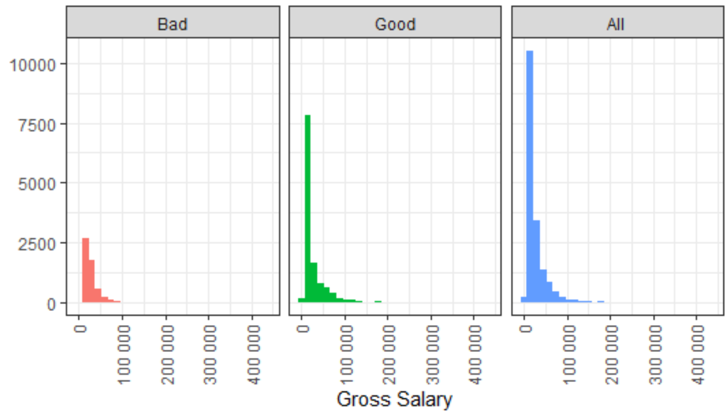

Figure 33: EDA: Gross Salary

5.1.13 Credit Ratio

Looking at credit ratio in Figure 34 below, we see that bad accounts reaches higher credit ratio compared to good accounts. Bad accounts have a mean of 0.162 with a variance of 0.00650 while good accounts have a mean of 0.147 and a variance of 0.00315. The higher mean of bad accounts is expected as those with higher utilization of credit lines are expected to have worse creditworthiness than those with lower utilization. Overall, we have a mean of 0.152 and a variance of 0.00424. Similarly, we must point out that bad accounts have about double the variance of good accounts. Hence, we need to inspect how credit ratio affects creditworthiness.

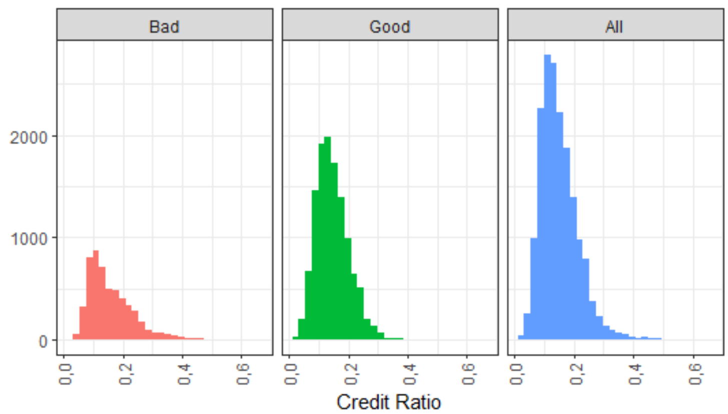

Figure 34: EDA: Credit Ratio

5.2 Single Factor Analysis

In order to perform single factor analysis, we split the accounts data into training and test datasets using the `scorecard::split_df()` function with seed 314. We then created logical binnings and obtained their weight of evidences from the training dataset using the function `scorecard::woebin()`. The Weight of Evidence (WOE) can be computed by

$$\text{WOE}_c = \ln\left[\frac{P(c|\text{Good})}{P(c|\text{Bad})}\right],$$

where c is a category. We also compared the results to the test dataset to confirm if the test dataset also follows the logical binnings. Furthermore, using the `scorecard::iv()` function, we can obtain the information value of each variable. The information value is given by

$$\text{IV} = \sum_{c=1}^{C} \text{WOE}_c \left[P(c|\text{Good}) - P(c|\text{Bad})\right],$$

where C is the number of categories under the variable. We have the following initial information value for each variable in Table 1. Since Civil Status and Account Number have an information value less than 0.02, we ignore these variables since they are generally unpredictive. Moreover, we can also omit the Sex variable to ensure equal credit opportunity for all genders.

Variable	Value
Amortization	1.27241006713
Months Loan	0.75954616090
Gross Salary	0.59671723317
House Type	0.51065953693
Education	0.37590948810
Credit Status	0.33361879804
Years Employed	0.22738912310
Dependents	0.19235570973
Purpose Loan	0.15486439763
Credit Ratio	0.12979943587
Sex	0.04485692631
Civil Status	0.00226511980
Account No	0.00008658149

Table 1: Initial Information Value

5.2.1 Dependents

In line with the EDA, Figure 35 indicates that accounts with more dependents are less likely to be bad accounts. However, we notice that clients with 0 dependent are more likely to be good client compared to clients with 1 to 3 dependents. We note that its information value is 0.1890 indicating a medium predictor. We also note that both training and test datasets follow the logical binnings.

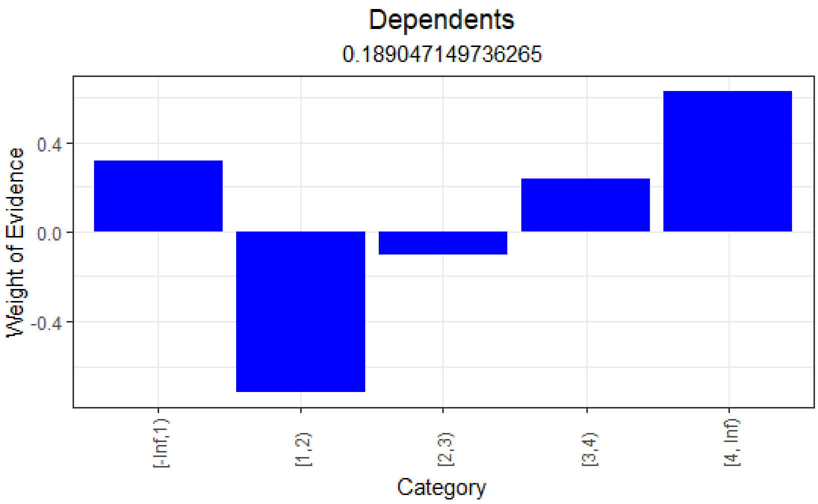

Figure 35: SFA: Dependents WOE

5.2.2 House Type

The WOEs in Figure 36 below agrees with the observations from the EDA. Owned houses were more likely to be good accounts while the other house types were more likely to be bad accounts. Missing info was also consider as category for this variable. Moreover, we have an information value of 0.5191 indicating a strong predictor. Moreover, both training and test datasetst follow the logical binnings.

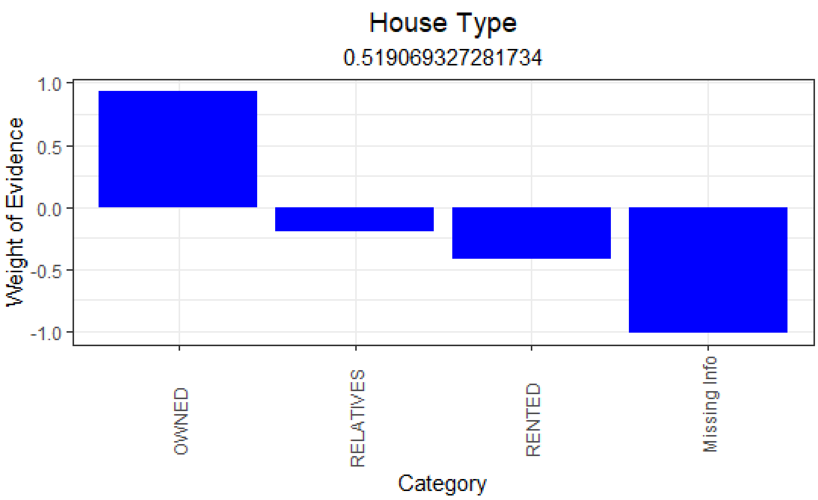

Figure 36: SFA: House Type WOE

5.2.3 Education

For education, we see in Figure 37 that accounts with college or vocational education have the highest WOE followed by high school education. Elementary education has the least WOE. This is in line with our expectations as the people who graduated from college and vocational courses are more likely to have a secure job and be able to pay back their loans. We have an information value of 0.3592 indicating a strong predictor. Furthermore, both training and test datasets agree on the logical binnings.

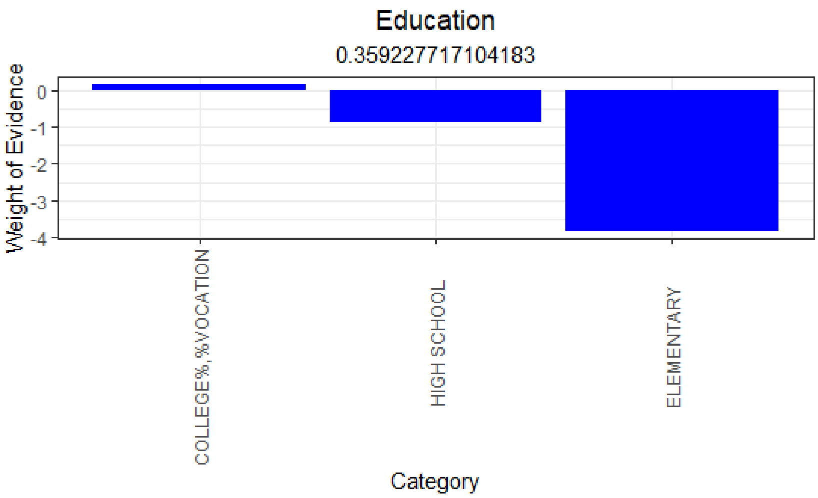

Figure 37: SFA: Education WOE

5.2.4 Years Employed

In Figure 38, we get a slightly different result from the EDA. Accounts with more years employed and with less years employed are more likely to be good accounts while accounts with 8 to 14 years employed are more likely to be bad accounts. However, its information value is 0.1766 indicating a medium predictor. Despite this, both training and test datasets agree on the logical binnings.

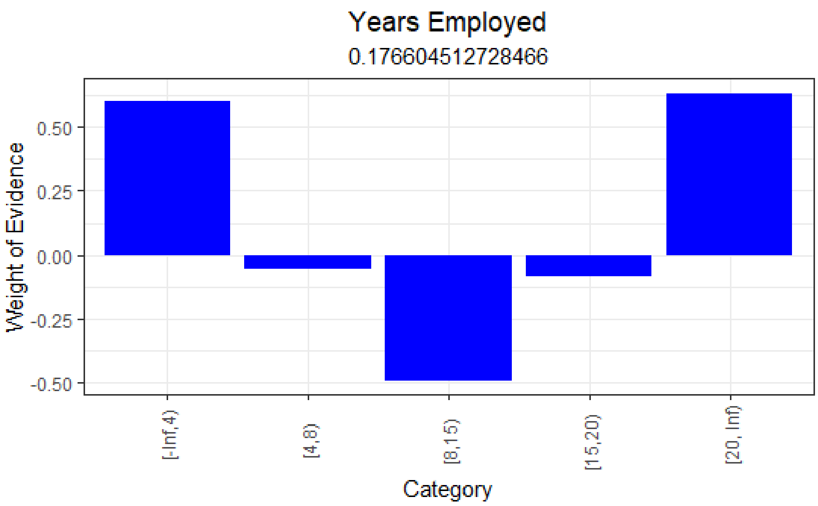

Figure 38: SFA: Years Employed WOE

5.2.5 Credit Status

Similar to the EDA, the WOEs in Figure 39 indicate that accounts with a credit status of paid-off are more likely to be good accounts followed by current accounts. Accounts with non-earning credit status are more likely to be bad accounts. We also arrive at an information value of 0.3579 indicating a strong predictor. Furthermore, both training and test datasets follow the logical binnings.

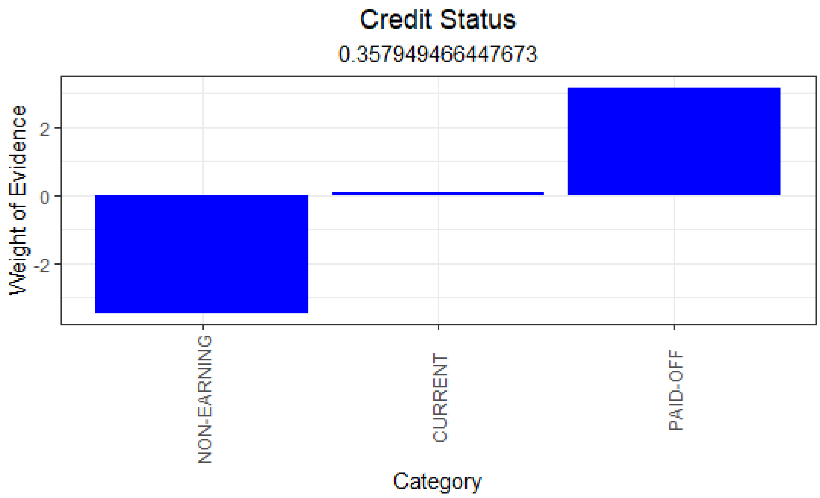

Figure 39: SFA: Credit Status WOE

5.2.6 Months Loan

The initial binning produced by R resulted in a binning that showed an increasing WOE as the number of months loan increases until 43 months. This is illogical as we know that the higher the number of months loan, the more likely the account will go bad. After rebinning, we see a similar observation in the EDA in Figure 40. Higher months loan indicate a higher likelihood of being a bad account since the higher the duration of the loan, the more likely it is to default. We also have an information value of 0.1478 indicating a medium predictor. Similarly, both training and test datasets follow the logical binnings.

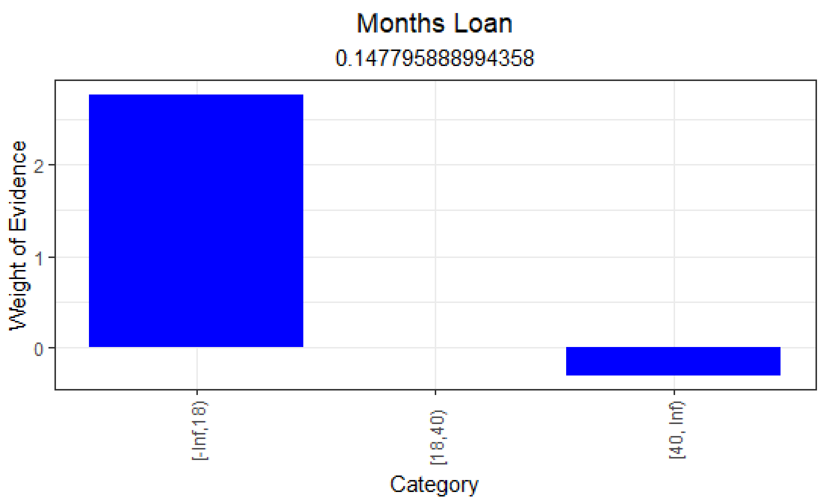

Figure 40: SFA: Months Loan WOE

5.2.7 Amortization

In the initial binning produced by R, the WOEs did not follow a logical trend. After rebinning, figure 41 below produces results similar to the observation from the EDA. Higher amortization indicates higher likelihood of being a bad account. We get an information value of 0.1700 indicating a medium predictor. Likewise, training and test datasets agree with the logical binnings.

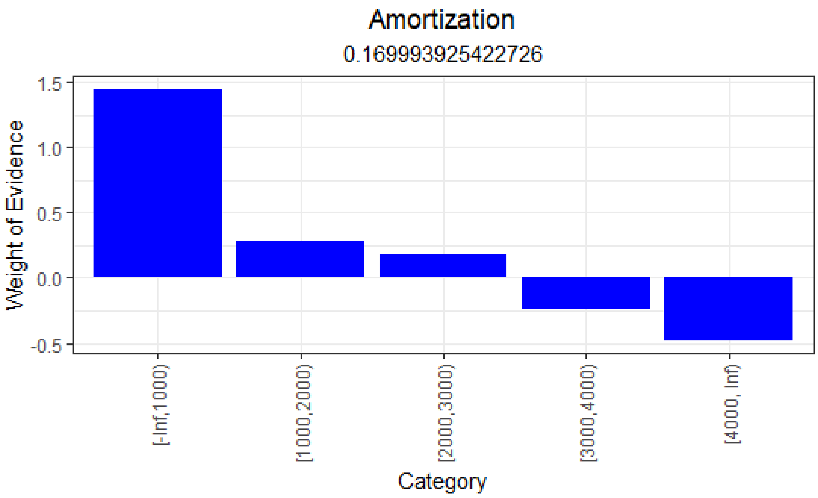

Figure 41: SFA: Amortization WOE

5.2.8 Purpose of Loan

The WOEs in Figure 42 supports the observation from the EDA that accounts with medical as the purpose are more likely to be bad accounts. In fact, we get an information value of 0.1509 indicating a medium predictor. Similarly, both types of datasets agree with the logical binnings.

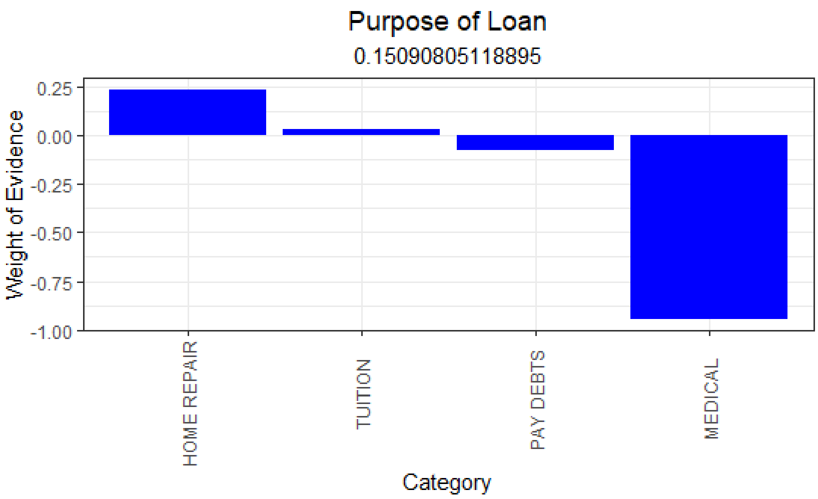

Figure 42: SFA: Purpose of Loan WOE

5.2.9 Gross Salary

For gross salary, we see a similar pattern in Figure 43 as the EDA. Higher gross salary points to a higher likelihood of being a good account. However, its information value is only 0.0502 indicating a weak predictor. Despite this, both types of datasets follows the logical binnings.

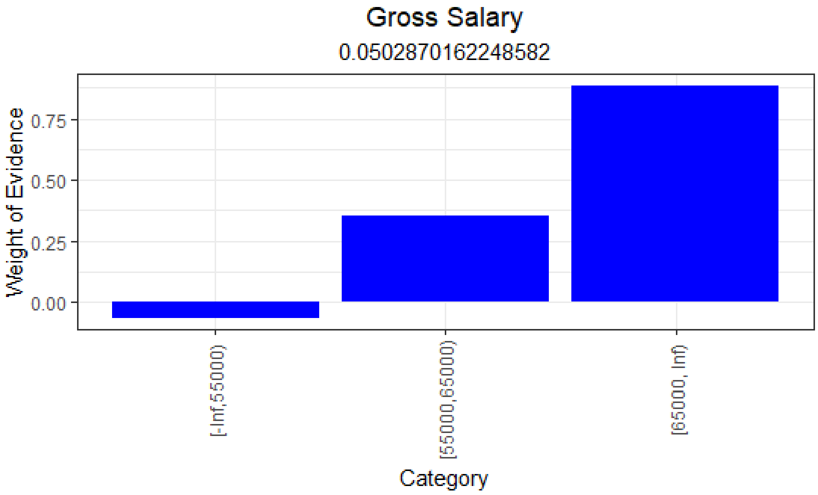

Figure 43: SFA: Gross Salary WOE

5.2.10 Credit Ratio

In line with the EDA, we see that accounts with higher credit ratio are more likely to be bad accounts in Figure 44. We arrive at an information value of 0.1038 indicating a medium predictor. Likewise, both training and test datasets agree on the logical binnings.

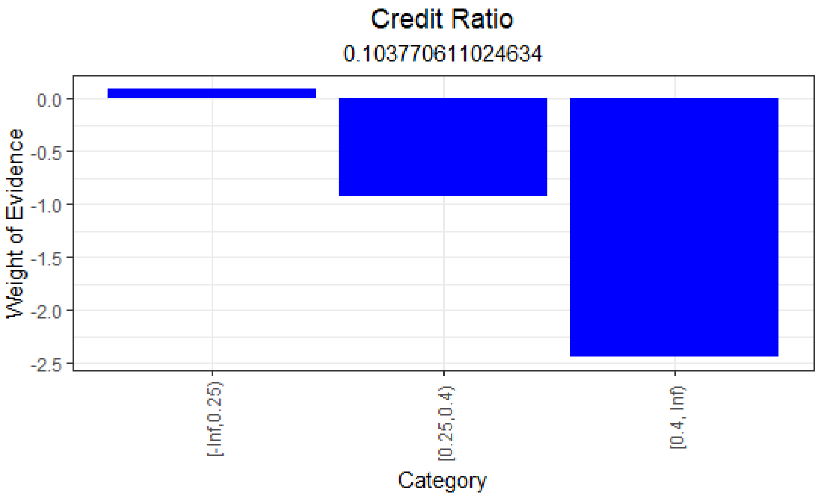

Figure 44: SFA: Credit Ratio WOE

5.3 Logistic Regression

After obtaining logical binnings and computing their weight of evidences, we performed a logistic regression on the variables. A logistic regression model is given by

$$\ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X^T\beta},$$

where $\boldsymbol{X}$ are the variables and $\boldsymbol{\beta}$ are the coefficients. In this case, $\boldsymbol{X}$ are the weight of evidences of each variable. For the logistic regression model, we create three models, as part of the variable selection procedure. We have one for forward, backward, and step-wise regression using AIC as the criteria. All three methods produced the same full model containing all 10 variables in the single factor analysis. Table 2 presents the summary of the result.

	Estimate	Standard Error	z-value
β_0	-0.81563	0.02296	-35.532*
`House_Type_woe`	-0.98551	0.03193	-30.865*
`Credit_Status_woe`	-1.37475	0.07670	-17.925*
`Education_woe`	-0.96043	0.05553	-17.296*
`Amortization_woe`	-1.29311	0.06570	-19.683*
`Months_Loan_woe`	-1.10827	0.10203	-10.862*
`Purpose_Loan_woe`	-0.82324	0.05591	-14.725*
`Gross_Salary_woe`	-1.19841	0.12043	-9.951*
`Credit_Ratio_woe`	-0.57304	0.07414	-7.729 *
`Dependents_woe`	-0.76388	0.05046	-15.138*
`Yrs_Employed_woe`	-0.82246	0.05338	-15.407*

* Significant variables at $\alpha = 0.05$.

Table 2: Summary of Logistic Regression

In equation, the fitted model is given by

$$\begin{aligned}\ln\left(\frac{\pi}{1-\pi}\right) = &- 0.81563 - 0.98551\ \texttt{House\_Type\_woe} - 1.37475\ \texttt{Credit\_Status\_woe}\\ &- 0.96043\ \texttt{Education\_woe} - 1.29311\ \texttt{Amortization\_woe}\\ &- 1.10827\ \texttt{Months\_Loan\_woe} - 0.82324\ \texttt{Purpose\_Loan\_woe}\\ &- 1.19841\ \texttt{Gross\_Salary\_woe} - 0.57304\ \texttt{Credit\_Ratio\_woe}\\ &- 0.76388\ \texttt{Dependents\_woe} - 0.82246\ \texttt{Yrs\_Employed\_woe}.\end{aligned}$$

After obtaining the model, we need to determine the optimal cutoff threshold. Two methods were considered for this. First, we find the optimal cutoff threshold that minimizes the distance of the receiver operating characteristic (ROC) curve to the point $(0, 1)$. More formally, we want to minimize the objective function $(1-\text{TPR})^2 + (\text{FPR})^2$, where TPR represents the true positive rate and FPR represents the false positive rate. This can be obtained using a ternary search algorithm given the bitonic nature of the distance of the ROC curve to the point $(0, 1)$. For the second method, we want to maximize the F1 score. This can be obtained using the function `scorecard::perf_eva()`. Using these two thresholds, we obtain the following metrics in Table 3 using the test dataset.

Metric	Method 1	Method 2
Threshold	0.32049329620880	0.32690000000000
Accuracy	0.77670000000000	0.78000000000000
Precision	0.61841070023603	0.62540192926045
Specificity	0.79361702127660	0.80170212765957
Sensitivity	0.73941674506115	0.73189087488241
F1 Score	0.67352185089974	0.67352185089974
AUC	0.83428594303557	0.83428594303557
AIC	12500	12500
Kolmogorov-Smirnov statistic	0.53410820439943	0.53410820439943
Gini	0.66857188607114	0.66857188607114
Somer's D	0.66796981645684	0.66796981645684
Hosmer-Lemeshow statistic	49.24213632842636	49.24213632842636
Hosmer-Lemeshow p-value	0.00000005713182	0.00000005713182

Table 3: Summary of Metrics

If TP is the true positives, TN is the true negatives, FP is the false positives, and FN is the false negatives, the metrics can be computed with the following:

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}\\
\text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}}\\
\text{Sensitivity} &= \frac{\text{TP}}{\text{TP}+\text{FN}}\\
\text{Specificity} &= \frac{\text{TN}}{\text{TN}+\text{FP}}\\
\text{F1 Score} &= \frac{2\times\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}}
\end{aligned}$$

If N_j is the number of observations, O_j is the number of bad accounts, and $\overline{\pi}_j$ is the average probability among bad accounts in the j^{th} decile,

$$\chi^2_{HL} = \sum_{j=1}^{n} \frac{(O_j - N_j\overline{\pi}_j)^2}{N_j\overline{\pi}_j(1-\overline{\pi}_j)}$$

If n_C is the number of concordant pairs and n_D is the number of discordant pairs,

$$\text{SD} = \frac{n_C - n_D}{T}$$

If F_B, F_G are the empirical cdf of bad accounts and good accounts, respectively

$$\text{KS} = \sup_z |F_B(z) - F_G(z)|$$

If AUC is the area under the ROC curve,

$$\text{AUC} = \frac{1+\text{Gini}}{2}$$

Since the objective of our model is to determine the bad accounts, our priority is to have a high sensitivity. We want to prioritize higher sensitivity to avoid giving out loans to people who are likely to default. Thus, we shall choose the threshold from the first method. Furthermore, Figure 45 shows the ROC curve, F1 test, KS test, and Gini coefficient of the model.

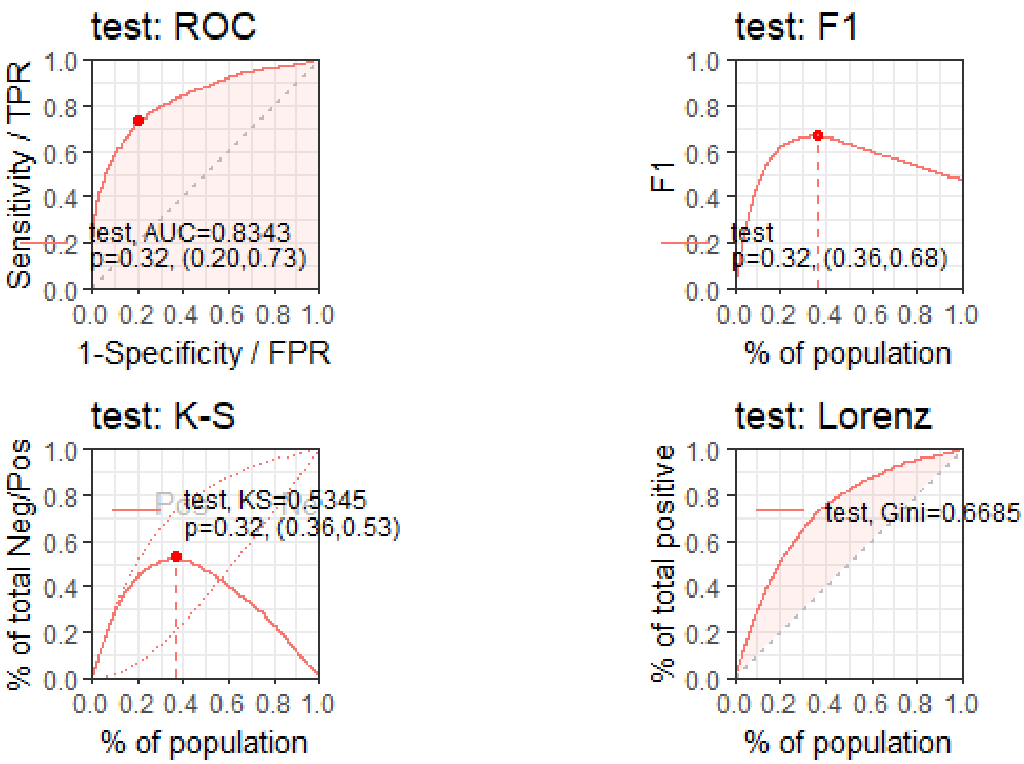

Figure 45: Model Results

Using the threshold of the first method, we obtain the following confusion matrix in Table 4.

		Predicted	
		0	1
Actual	0	1865	485
	1	277	786

Table 4: Confusion Matrix

We also check for multicolinearity with the function `scorecard::vif()` to obtain the Variance Inflation Factor. The variance inflation factor can be computed using

$$\mathrm{VIF}_k = \frac{1}{1 - \widehat{R_k}^2},$$

where $\widehat{R_k}$ is the adjusted R-squared for the regression of X_k on the other covariates. Table 5 in the following page shows the variance inflation factor for the variables. Since all values are near 1, we can conclude that there is no multicolinearity.

Variable	VIF
`Amortization_woe`	1.206500
`Credit_Ratio_woe`	1.076639
`Credit_Status_woe`	1.192904
`Dependents_woe`	1.005624
`Education_woe`	1.009878
`Gross_Salary_woe`	1.145763
`House_Type_woe`	1.019393
`Months_Loan_woe`	1.188534
`Purpose_Loan_woe`	1.004388
`Yrs_Employed_woe`	1.007921

Table 5: Variance Inflation Factor

5.4 Scorecard Development

To convert our logistic model into a scorecard, we will use the following formula

$$\begin{aligned}\text{Score}_j &= \text{Offset} + \text{Factor} \times \ln\left(\frac{1-\pi_j}{\pi_j}\right)\\ &= \underbrace{[\text{Offset} - \text{Factor} \times \beta_0]}_{\text{Base Score}} + \sum_{m=1}^{k} \underbrace{-\text{Factor} \times \beta_m \times \text{WOE}_{m,j}}_{\text{Points}_{m,j}}.\end{aligned}$$

Given two points (odds_s, s) and $(2 \times \text{odds}_s, s + p_2)$, we can obtain the following

$$\text{Factor} = \frac{p_2}{\ln 2}$$
$$\text{Offset} = s - \text{Factor} \times \ln(\text{odds}_s).$$

For our model, we shall use $s = 600$, $\text{odds}_s = 50$, and $p_2 = 20$. Using these, we get

$$\text{Factor} = 28.8539$$
$$\text{Offset} = 487.1229$$
$$\text{Cutoff} = 508.8067.$$

From these values, we obtain the final credit scorecard model in Table 6 in the following page.

Variable	Bin	Points
Amortization	$(-\infty, 1000)$	54
	$[1000, 2000)$	11
	$[2000, 3000)$	7
	$[3000, 4000)$	-9
	$[4000, \infty)$	-18
Credit Ratio	$(-\infty, 0.25)$	1
	$[0.25, 0.4)$	-15
	$[0.4, \infty)$	-40
Credit Status	Current	2
	Non-earning	-138
	Paid-off	125
Dependents	$(-\infty, 1)$	7
	$[1, 2)$	-16
	$[2, 3)$	-2
	$[3, 4)$	5
	$[4, \infty)$	14
Education	College, Vocation	4
	Elementary	-107
	High school	-24
Gross Salary	$(-\infty, 55000)$	-2
	$[55000, 65000)$	12
	$[65000, \infty)$	31
House Type	Missing Info	-29
	Owned	26
	Relatives	-6
	Rented	-12
Months Loan	$(-\infty, 18)$	88
	$[18, 40)$	0
	$[40, \infty)$	-10
Purpose Loan	Home repair	6
	Medical	-22
	Pay debts	-2
	Tuition	1
Years Employed	$(-\infty, 4)$	14
	$[4, 8)$	-1
	$[8, 15)$	-12
	$[15, 20)$	-2
	$[20, \infty)$	15
Base Points		511

Table 6: Credit Scorecard - Essential

6 The Segmented Credit Scorecard - Non-Essentials

For the second credit scorecard model, we only consider the accounts where the purpose of loan is non-essential. These are the accounts that took out a loan for either leisure, travel, purchase of appliance, or to purchase an auto. Among the 23,321 accounts, 5,921 or around 25.39% were for purposes that were considered non-essential. Of the 5,921 accounts, 4,820 were good accounts representing 81.40% of the total, while 1,101 were bad accounts representing 18.60% of the total. A large percentage of these loans were for leisure purposes, with 5,038 of the 5,921 being for this purpose. For the rest, 301 were for appliance, 332 were auto loans, and 250 were for travel.

6.1 Exploratory Data Analysis

We first perform an exploratory data analysis on the entire segmented non-essential dataset. We compared the distribution of each variable across the good accounts, the bad accounts, and the overall population in order to gain insights on the correlation of each variable with creditworthiness.

6.1.1 Account Number

First, we take a look at the account numbers. From Figure 46, we see that the good accounts and bad accounts are roughly equally distributed across the account numbers. No trend can be identified from the graphs. The bad accounts have mean 11143.39 and variance 45048680. The good accounts have mean 111721.92 and variance 44482492. All accounts have mean 11614.34 and variance 44630854. The mean and variance of the good, the bad, and all of the accounts, do not seem much to vary from each other. This confirms the initial observation that the good and bad accounts are roughly equally distributed across the account numbers. This is as expected as account numbers should generally not affect creditworthiness. The insights found in the EDA of this segment of the datset concur with the insights found in the EDA of the entire datset. Thus, there might not be a strong correlation between the account number and the creditworthiness.

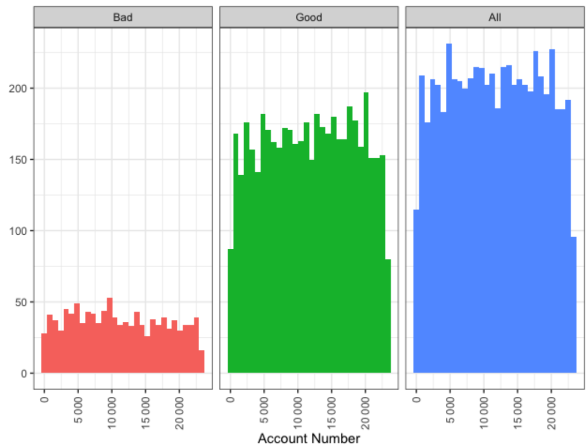

Figure 46: EDA: Account Number

6.1.2 Sex

From Figure 47, we can see that much more male who take loans as compared to female. On the other hand, there is only a small difference number of male and female who were classified under bad creditworthiness. Under bad creditworthiness, 461 were female while 640 were male. On the other hand, 3456 male and only 1364 female with good creditworthiness. Thus, looking in terms of proportion, we can say that in general, female accounts are more likely to default as compared to male accounts. Sex might have a strong correlation with creditworthiness. This is similar to the findings in the EDA of the entire dataset. However, it is important to consider political factors and to avoid issues of discrimination, we may elect not to include this in the variables to be used in the model.

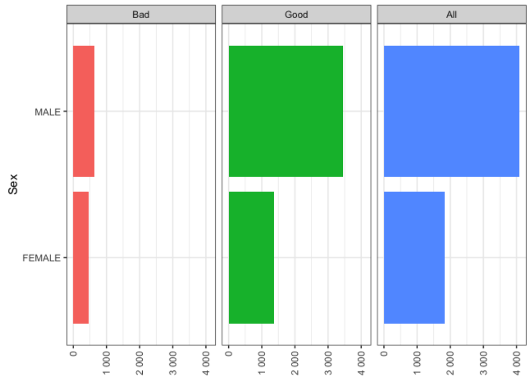

Figure 47: EDA: Sex

6.1.3 Dependents

From Figure 48, we see that a high percentage of the accounts with 1 dependent are good accounts. The mean number of dependents of those with bad creditworthiness is 1.319709 with a variance of 1.543148. On the other hand, those with good creditworthiness has an average of 1.231120 with variance 2.168610. The mean dependents for all accounts is 1.247593 with variance 2.053214. This shows that the average number of dependents of those with bad creditworthiness is slightly higher than those with good creditworthiness. From this, we can infer that less dependents could correlate to better creditworthiness. This is opposite that of the results found in the EDA of the entire dataset. We do take note however, that among the few accounts that have more than 5 dependents, most are considered to be good accounts.

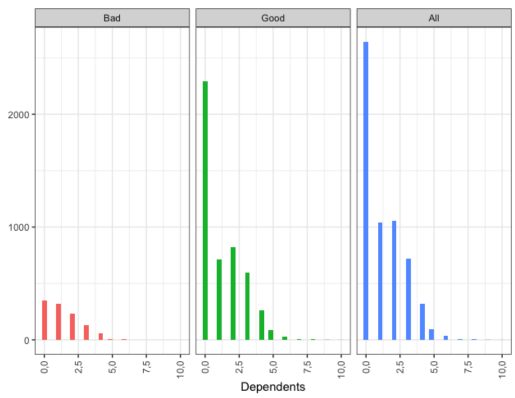

Figure 48: EDA: Dependents

6.1.4 Civil Status

Looking at Figure 49, we see that the distribution according to civil status of the good and bad accounts are roughly the same. Most of those who have loans are married, followed by those who are single, which is roughly 2/3 of the number of those who are married. The number of separated and widowed is very small compared to the total. Thus, we can infer that civil status may have a weak correlation with creditworthiness at best.

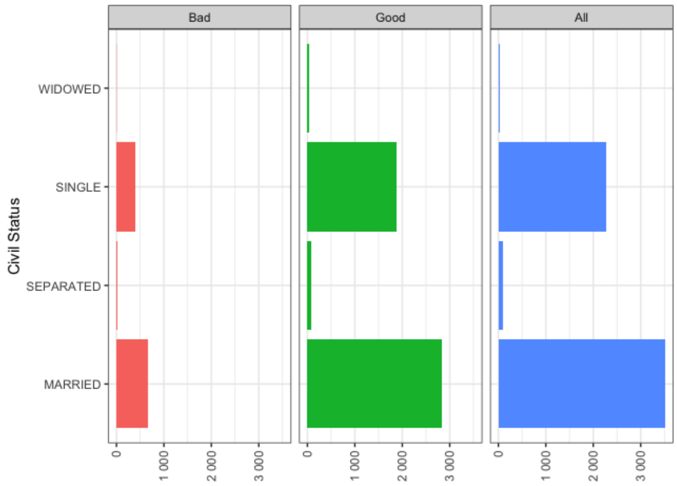

Figure 49: EDA: Civil Status

6.1.5 House Type

Looking at the graph in Figure 50, we can see in terms of proportion, those who owned their house are a much smaller percentage of the people with bad creditworthiness as compared to the other house ownership categories. The distribution did not seem to differ much in terms of the good and bad for the other categories. Thus, we can infer that from the house type, we can say that owning a house can be a good indicator of creditworthiness.

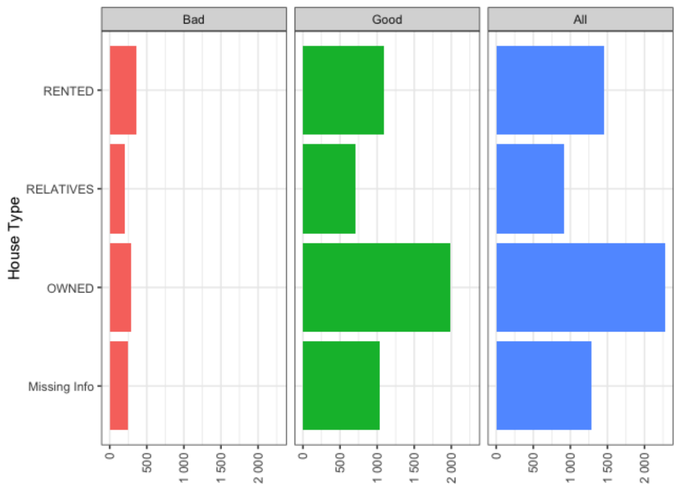

Figure 50: EDA: House Type

6.1.6 Education

In Figure 51, we see that most of those with elementary as their highest educational attainment have bad creditworthiness. Of the 69 with elementary as the highest educational attainment, 68 had bad creditworthiness. Moreover, those with highschool as the highest educational attainment, we also see that most of them have bad creditworthiness, with 80 out of the 129 having bad creditworthiness. On the other hand, those with college education generally have a large portion with good creditworthiness.

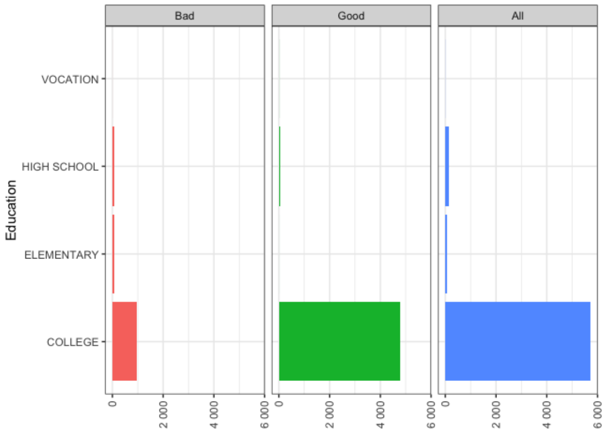

Figure 51: EDA: Education

6.1.7 Years Employed

From a first look at Figure 52, we see that those with bad creditworthiness have a higher average number of years of employment as compared to those with good creditworthiness. The bad accounts have an average of 8.499546 years of employment with a variance of 32.84477 while the good accounts have an average of 7.345228 with a variance of 41.76832. The average years of employments of all accounts is 7.559872. The result seem to be counterintuitive since intuition says the good accounts should generally have more years of employment but this can partially be due to the high variance. Moreover, we do take note that those with very few years of employment and those with very high years of employment, as compared to the entire set of accounts, have a much greater percentage of accounts that are considered to be good.

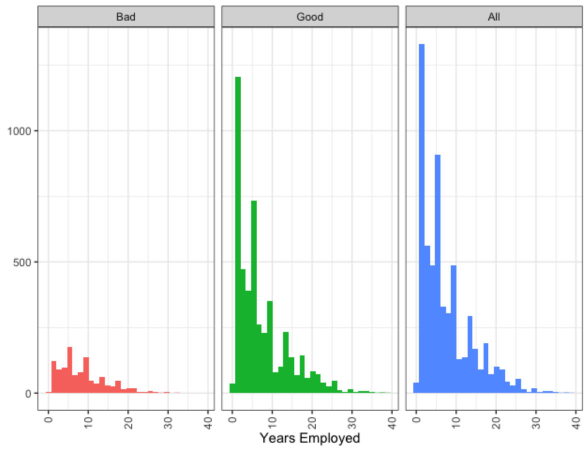

Figure 52: EDA: Years Employed

6.1.8 Credit Status

From Figure 53 below, we see that almost all accounts that are paid off have good creditworthiness while those that are non-earning are mostly bad accounts. For the current accounts, a large percentage are good accounts and a small percentage are bad acounts. Thus, we can see that credit status can be a good indicator for creditworthiness.

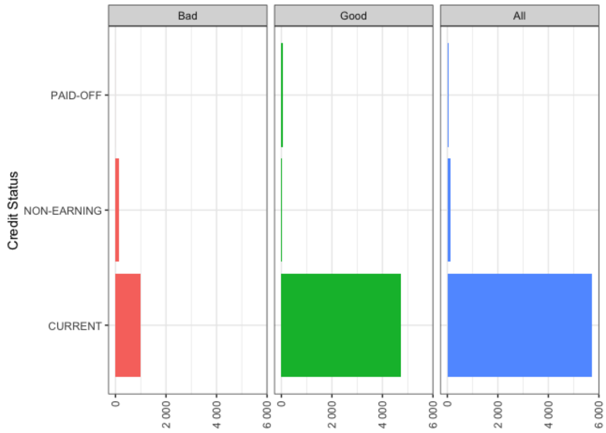

Figure 53: EDA: Credit Status

6.1.9 Months Loan

In Figure 54, we can see that accounts that have longer duration have a larger percentage of bad accounts as compared to those with shorter duration. The mean duration of all accounts is 28.29454 with a variance of 107.1612. The mean duration of bad accounts is higher, at 30.60581. On the other hand, the mean duration of the good accounts is 27.76660. Thus, the monhts loan variable seems to have a correlation with the creditworthiness of the account. Moreover, for those with very high loan duration, we also see a spike in the percentage of the accounts that are considered to be bad. Thus, we can infer that the longer the loan, the higher the chance for it to be bad. This is consistent with the idea of increasing risk as time increases.

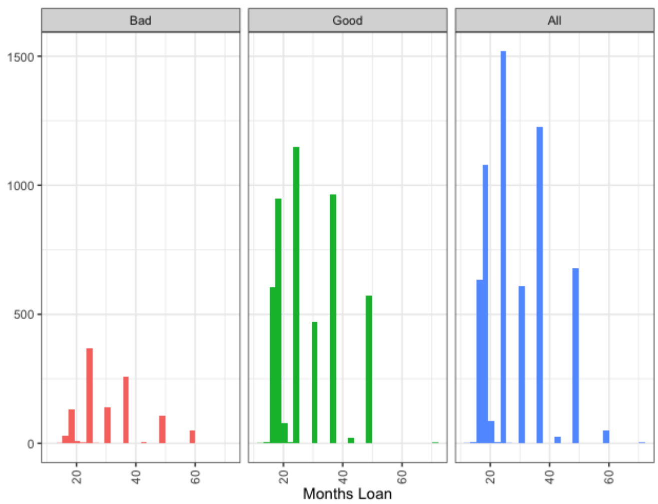

Figure 54: EDA: Months Loan

6.1.10 Amortization

We see in Figure 55 that the average amortization for all accounts is 6002.051 with a variance of 75394317. The average amortization for the good accounts is 6067.163 with a variance of 72300829. Meanwhile, the average for bad accounts is much lower at 5717.000 with a variance of 88915245.There is a difference between the means however, the large variance makes the difference small in comparison.

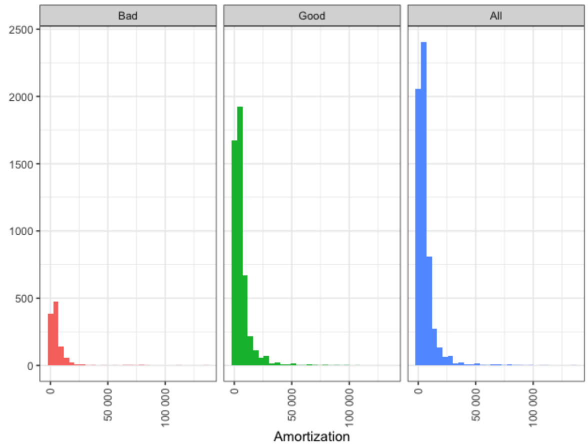

Figure 55: EDA: Amortization

6.1.11 Purpose of Loan

We see in Figure 56 that the spread across the different purposes seem to be the same for the good and bad accounts. Most of the loans taken out under both good and bad creditworthiness, are for leisure, with the next purpose with most loans being auto loans, followed by appliance and finally travel. This indicates that the purpose may have weak correlation with the creditworthiness of an account at best.

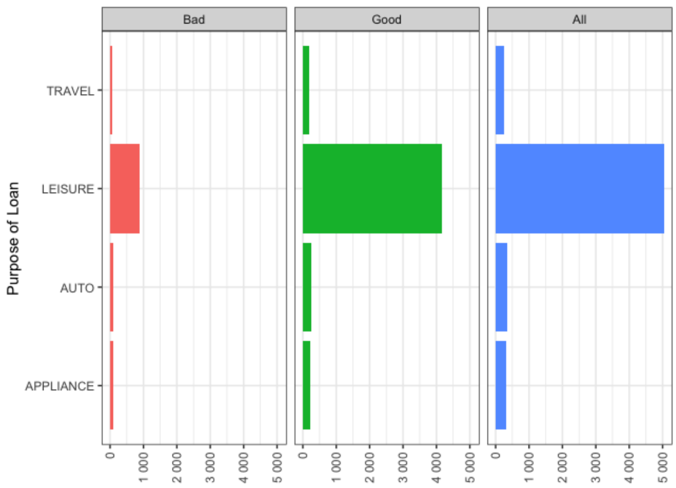

Figure 56: EDA: Purpose of Loan

6.1.12 Gross Salary

As expected, we see in Figure 57 that those with higher gross salary generally have better creditworthiness. The mean gross salary for all accounts is 39937.89 with a variance of 2186072491. The average gross salary of those with bad accounts is 32258.63 with a variance of 1514207713. On the other have, the good accounts have a much higher average gross salary which is at 41692.01 with variance 2323337351. Thus, the gross salary seem to be a good indicator of the creditworthiness of an account.

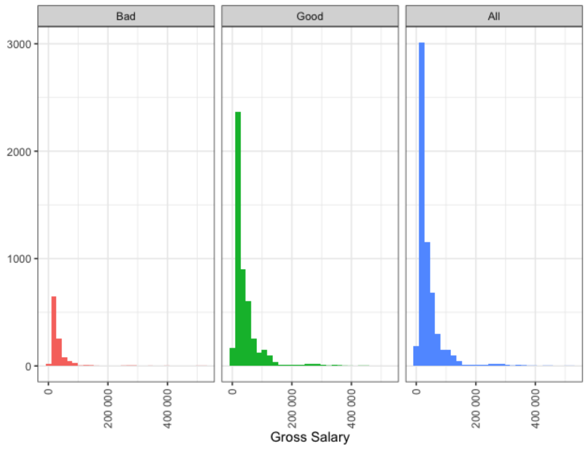

Figure 57: EDA: Gross Salary

6.1.13 Credit Ratio

In Figure 58, we see that most of the accounts have a credit ratio lower than 0.25. However, for those with much higher credit ratios, we can see that a larger proportion of these are bad accounts. Furthermore, we see that the bad account on average have a higher credit ratio. The average credit ratio of all accounts is 0.1500229 with a variance of 0.004573599. The bad accounts have a mean credit ratio of 0.1637525 and variance 0.005848038 as compared to the good accounts with an average credit ratio of 0.1468868 and variance 0.004230736. This supports our earlier observation. Thus, it is necessary to examine how creditworthiness is influenced by the credit ratio.

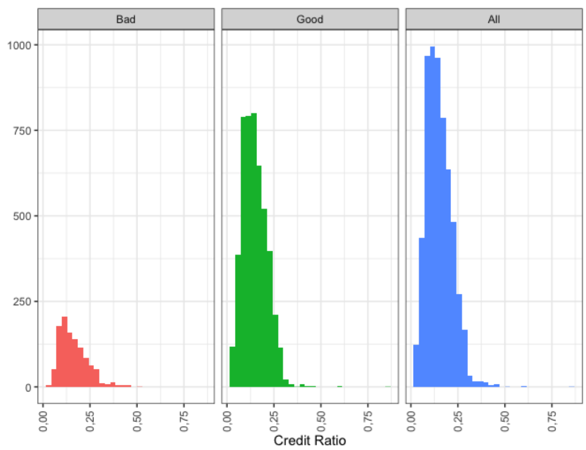

Figure 58: EDA: Credit Ratio

6.2 Single Factor Analysis

In order to perform single factor analysis, we first split the accounts into training and test datasets using the `scorecard::split_df()` function with seed 314. Binnings were then created using the function `scorecard::woebin()`. The WOEs of each binning were then taken. As previously stated, the WOE can be computed by

$$\text{WOE}_c = \ln\left[\frac{P(c|\text{Good})}{P(c|\text{Bad})}\right].$$

The WOEs obtained is then checked for logical consistency. If the initial binning obtained was illogical, rebinning was then done to create a binning with a much more logical trend. We also compared the results to the test dataset to confirm if the test dataset also follows the logical binnings. Furthermore, using the `scorecard::iv()` function, the information value of each variable was obtained. The information value is given by

$$\text{IV} = \sum_{c=1}^{C} \text{WOE}_c \left[P(c|\text{Good}) - P(c|\text{Bad})\right].$$

We have the following initial information value for each variable in Table 7. From the results, we can ignore Civil Status and Account Number since they have an information value less than 0.02 indicating that they are generally unpredictive. Purpose of the loan, Credit Ratio, and Sex are weak predictors. House type, Dependents, and Employment Years are medium predictors while the rest of the variables, namely- Gross salary, Education, Months Loan, Credit Status, and Amortization, are strong predictors. Moreover, we also omit the Sex variable to avoid gender discrimination and ensure equal credit opportunity.

Variable	Value
Amortization	1.0916041061
Credit Status	0.5108041544
Months Loan	0.5081837713
Education	0.4903447271
Gross Salary	0.4342807702
Years Employed	0.2200084267
Dependents	0.1806803399
House Type	0.1142224312
Sex	0.0816535192
Credit Ratio	0.0683112322
Purpose Loan	0.0250900596
Civil Status	0.0022812221
Account No	0.0000611598

Table 7: Initial Information Value

6.2.1 Dependents

As can be seen in Figure 59 accounts with 1 or less dependents are less likely to be bad accounts. Moreover, we also noted from the EDA that accounts with high number of dependents have a high percentage of good accounts. The information value from this is obtained to be 0.1282, indicating that number of Dependents is a medium predictor. We also note that both training and test datasets follow the same logical binnings.

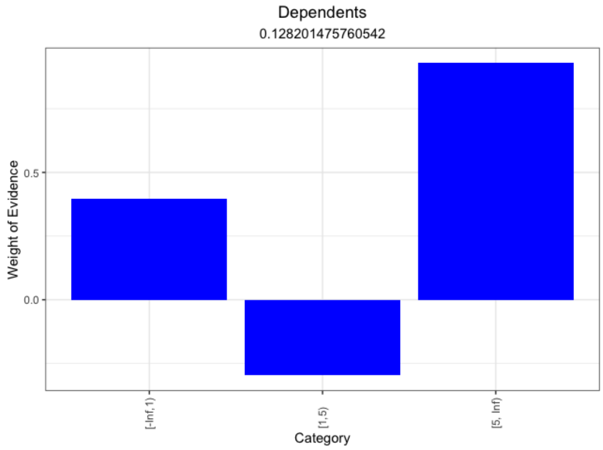

Figure 59: SFA: Dependents WOE

6.2.2 House Type

The WOEs in Figure 60 agrees with the observations from the EDA. Owned houses were more likely to be good accounts while the other house types were more likely to be bad accounts. Rented houses and those living with relatives were binned together. Missing info was consider as one bin for this binning. An information value of 0.2200 was obtained indicating that house type is a medium indicator. Furthermore, we note that the training and test datasets followed the same logical binnings.

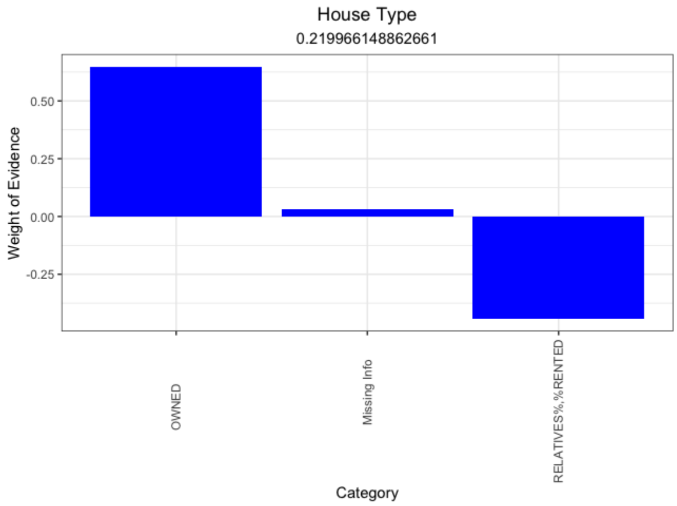

Figure 60: SFA: House Type WOE

6.2.3 Education

For education, we see in Figure 61 that accounts with college or vocational education were binned together since only the train set has accounts that had vocational education and the WOE obtained from the initial binning showed a close WOEs for these two binnings. The bin of college and vocational have the highest WOE followed by high school education. This is as expected as the people who graduated from college and vocational courses are more likely to have better opportunities and thus, be able to pay back their loans. Elementary education has the least WOE. An information value of 0.4624 indicating a strong predictor. Furthermore, both training and test datasets agree on the logical binnings.

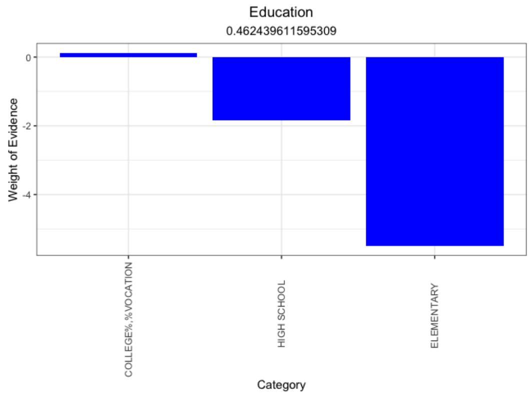

Figure 61: SFA: Education WOE

6.2.4 Years Employed

The initial binning by R produced a binning that showed a decreasing WOE as the years of employment increased. This is illogical as we know that the higher the number of years of employment a person has, the more stable their financial standing is and the more likely that their account will be good. Moreover, from the EDA, we saw that those with very few years of employment and very high years of employment have a high good rate. After rebinning, we obtained a binning that produced a logical result that was in concurrence with the findings of EDA as shown in Figure 62. The information value of the binning obtained is 0.1512, making it a medium predictor variable. The same trend was found on both the training, with the WOE starting out high and decreasing until the [6,11) bin. Afterwards, the WOE increases with the increase number of years employed.

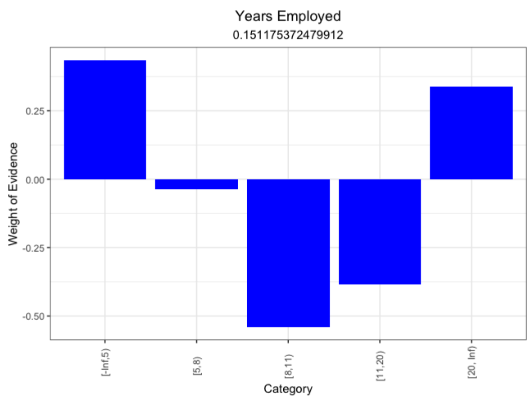

Figure 62: SFA: Years Employed WOE

6.2.5 Credit Status

As can be seen in Figure 63, the resulting WOEs from the initial optimal binning obtained are logical. The non-earning credit status has the lowest WOE, suggesting that this status is and indicator that the account is more likely to be a bad account. The Paid-Off status had the highest WOE. This results is consistent with the observations from the EDA performed at the earlier part of this paper. The IV is also high at 0.5227, which means that the Credit Status is a strong predictor of creditworthiness. The logic was also consistent in both the train and test sets.

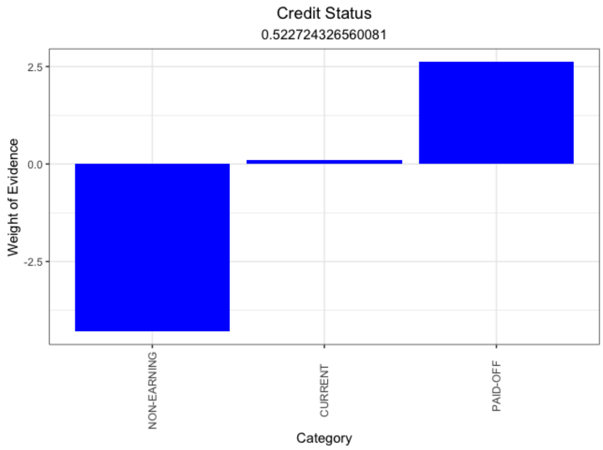

Figure 63: SFA: Credit Status WOE

6.2.6 Months Loan

The initial binning made by R did not follow a logical trend for the WOEs. Rebinning was done and in the obtained logical binning, we see in Figure 64 that the higher the months of the loan, the higher the likelihood of it being a bad loan. This is in line with the observation in the EDA that the higher the duration of the loan, the more risky it is. The information value of this model is 0.2167, indicating that it is a medium predictor. The logic is consistent in both the train and the test sets.

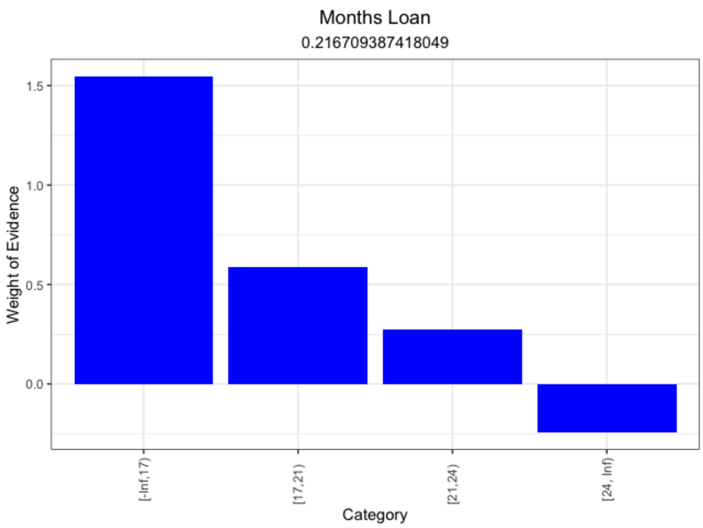

Figure 64: SFA: Months Loan WOE

6.2.7 Amortization

In the initial binning, the WOEs did not follow a logical trend. As stated in the EDA earlier, the higher the amortization, the more likely it is to become a bad account. After rebinning, we got a downward trend for the WOEs as the amortization amount increased. The WOEs and the binning can be seen in Figure 65. The same logical trend was seen in both the test and the training datasets. However, the information value decreased to 0.01917, making this a generally unpredictive variable.

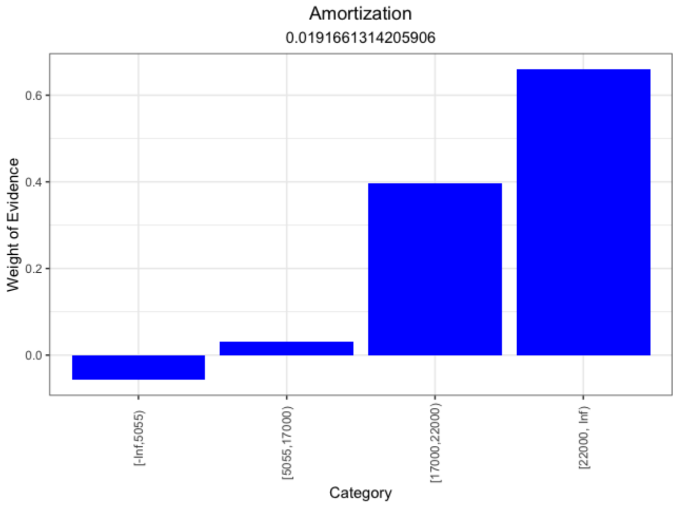

Figure 65: SFA: Amortization WOE

6.2.8 Purpose of Loan

For the purpose of the loan, the initial optimal binning shown in Figure 66 resulted in having Leisure and Travel as one bin, which makes sense since they are very similar categories. Leisure and Travel had the highest WOE, while Auto had the 2nd highest and Appliance had the lowest WOE among the 3. An IV of 0.0207 was obtained, indicating that purpose of loan is a weak predictor. The trend of the WOEs were the same for both the test and training dataset.

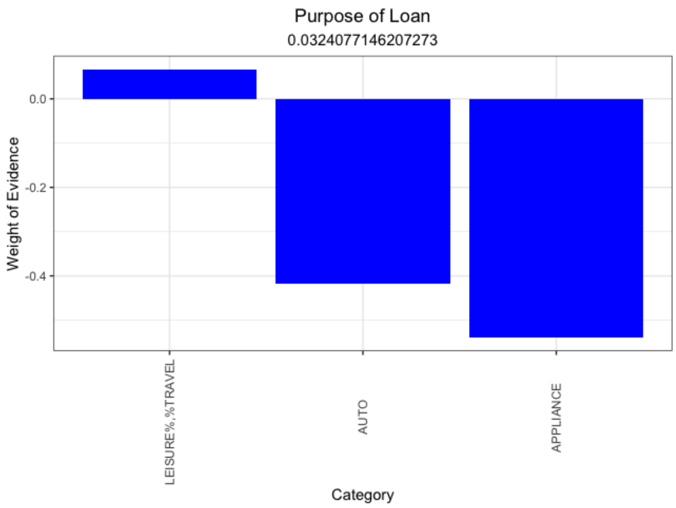

Figure 66: SFA: Purpose of Loan WOE

6.2.9 Gross Salary

For gross salary, as can be seen in Figure 67 a similar pattern to that observed in the earlier EDA was obtained. Higher gross salary points to a higher likelihood of being a good account. The information value of this binning was 0.1087, indicating that Gross Salary is a medium predictor.

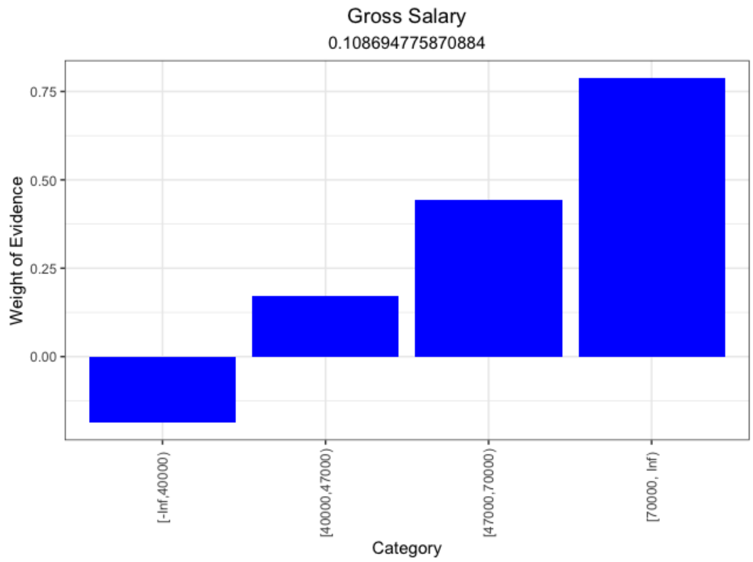

Figure 67: SFA: Gross Salary WOE

6.2.10 Credit Ratio

For the Credit ratio, the initial optimal binning produced a logical trend of the WOEs, as can be seen in Figure 68, consistent with the observation from the EDA that the higher the credit ratio, the more likely an account is to be bad. An IV of 0.1404 was obtained indicating that this is a medium predictor. The same logical trend in the WOEs of the binnings was seen in both the test and train sets.

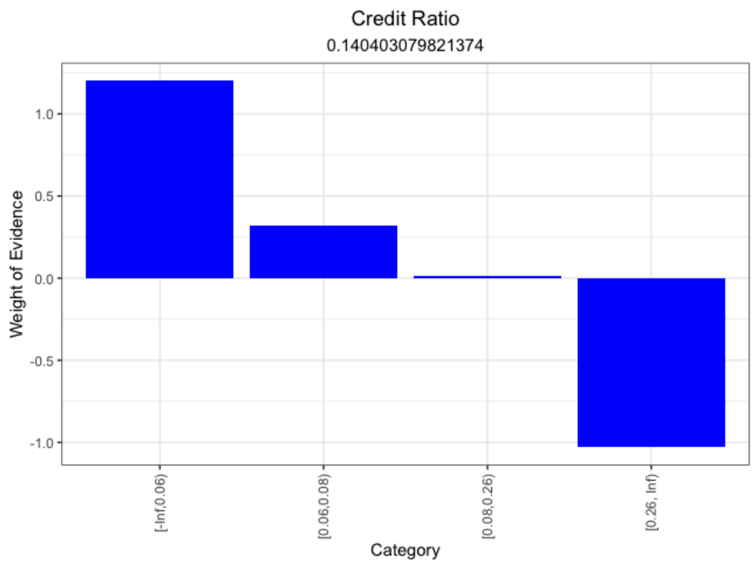

Figure 68: SFA: Credit Ratio WOE

6.3 Logistic Regression

After obtaining logical binnings and computing their weight of evidences, we want to perform a logistic regression on the variables. As stated in the first model, a logistic regression model is given by

$$\ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X^T\beta},$$

where $\boldsymbol{X}$ are the variables and $\boldsymbol{\beta}$ are the coefficients. To reiterate, $\boldsymbol{X}$ are the weight of evidences of each variable. For the logistic regression model, we create three models, as part of the variable selection procedure. First, we have the model created by forward regression, starting from the null model adding variables and the models adds variables one by one using AIC as the criteria. Another model is created with backward regression, starting with a model that included all the variables, then the variables are removed one by one using AIC as the criteria. Finally, we have step-wise regression, which is a combination of the forward and backward regression. Starting with the null model, at each step the program either adds or removes one variable also using AIC as the criteria. In this case, all three methods produced the same model containing all variables except Amortization. Table 8 presents the estimate, standard error and z-value for each coefficient.

	Estimate	Standard Error	z-value
β_0	-1.48996	0.04473	-33.310*
`Credit_Status_woe`	-1.13351	0.09900	-11.450*
`Education_woe`	-0.84569	0.09039	-9.356*
`Months_Loan_woe`	-0.92251	0.13901	-6.637*
`Credit_Ratio_woe`	-1.05160	0.12279	-8.564*
`Yrs_Employed_woe`	-0.89771	0.11167	-8.039*
`House_Type_woe`	-0.61975	0.08996	-6.889*
`Dependents_woe`	-0.84350	0.12347	-6.831*
`Purpose_Loan_woe`	-0.53101	0.22560	-2.354*
`Gross_Salary_woe`	-0.35741	0.16685	-2.142*

* Significant variables at $\alpha = 0.05$.

Table 8: Summary of Logistic Regression

In equation, the fitted model is given by

$$\begin{aligned}\ln\left(\frac{\pi}{1-\pi}\right) = &- 1.48996 - 1.13351\ \texttt{Credit\_Status\_woe} - 0.84569\ \texttt{Education\_woe}\\ &- 0.92251\ \texttt{Months\_Loan\_woe} - 1.05160\ \texttt{Credit\_Ratio\_woe}\\ &- 0.89771\ \texttt{Yrs\_Employed\_woe} - 0.61975\ \texttt{House\_Type\_woe}\\ &- 0.84350\ \texttt{Dependents\_woe} - 0.53101\ \texttt{Purpose\_Loan\_woe}\\ &- 0.35741\ \texttt{Gross\_Salary\_woe}.\end{aligned}$$

After the model was obtained, an optimal cutoff threshold is obtained. Two different methods were used to determine an appropriate cutoff. First, an optimal cutoff threshold that minimizes the distance of the ROC curve to the point $(0, 1)$. That is, we find the cutoff that minimizes the function $(1-\text{TPR})^2+(\text{FPR})^2$, where TPR represents the true positive rate and FPR represents the false positive rate. This can be obtained using a ternary search algorithm given the bitonic nature of the distance of the ROC curve to the point $(0, 1)$. On the other hand, the second method maximizes the F1 score. This is obtained using the function `scorecard::perf_eva()`.

To reiterate what was said in the first model, the metrics for evaluation of the threshold are computed as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \\
\text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\
\text{Sensitivity} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\
\text{Specificity} &= \frac{\text{TN}}{\text{TN}+\text{FP}} \\
\text{F1 Score} &= \frac{2\times\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}}
\end{aligned}$$

If N_j is the number of observations, O_j is the number of bad accounts, and $\overline{\pi}_j$ is the average probability among bad accounts in the j^{th} decile,

$$\chi^2_{HL} = \sum_{j=1}^{n} \frac{(O_j - N_j\overline{\pi}_j)^2}{N_j\overline{\pi}_j(1-\overline{\pi}_j)}$$

If n_C is the number of concordant pairs and n_D is the number of discordant pairs,

$$\text{SD} = \frac{n_C - n_D}{T}$$

If F_B, F_G are the empirical cdf of bad accounts and good accounts, respectively

$$\text{KS} = \sup_z |F_B(z) - F_G(z)|$$

If AUC is the area under the ROC curve,

$$\text{AUC} = \frac{1+\text{Gini}}{2}$$

A summary of the results using these two different thresholds is presented in Table 9 below.

Metric	Method 1	Method 2
Threshold	0.1889476	0.2122000
Accuracy	0.7309000	0.7581000
Precision	0.3712737	0.3993711
Specificity	0.7494600	0.7937365
Sensitivity	0.6492891	0.6018957
F1 Score	0.4724138	0.4801512
AUC	0.7669690	0.7669690
AIC	3637	3637
Kolmogorov-Smirnov statistic	0.4048089	0.4048089
Gini	0.5339379	0.5339379
Somer's D	0.5252372	0.5252372
Hosmer-Lemeshow statistic	9.0611632	9.0611632
Hosmer-Lemeshow p-value	0.3371626	0.3371626

Table 9: Summary of Metrics

In a credit scorecard, the main objective is to be able to manage risks by correctly identifying those that would become bad accounts, we prioritize having a high sensitivity for our model. Thus, we see that the threshold from method 1, which has a sensitivity of 0.6493 is chosen. The accuracy of this model since the objective of our model is to determine the bad accounts, our priority is to have a high sensitivity. Thus, we shall choose the threshold from the first method. Furthermore, Figure 69 on the following page shows the ROC curve, F1 test, KS test, and Gini coefficient of the model.

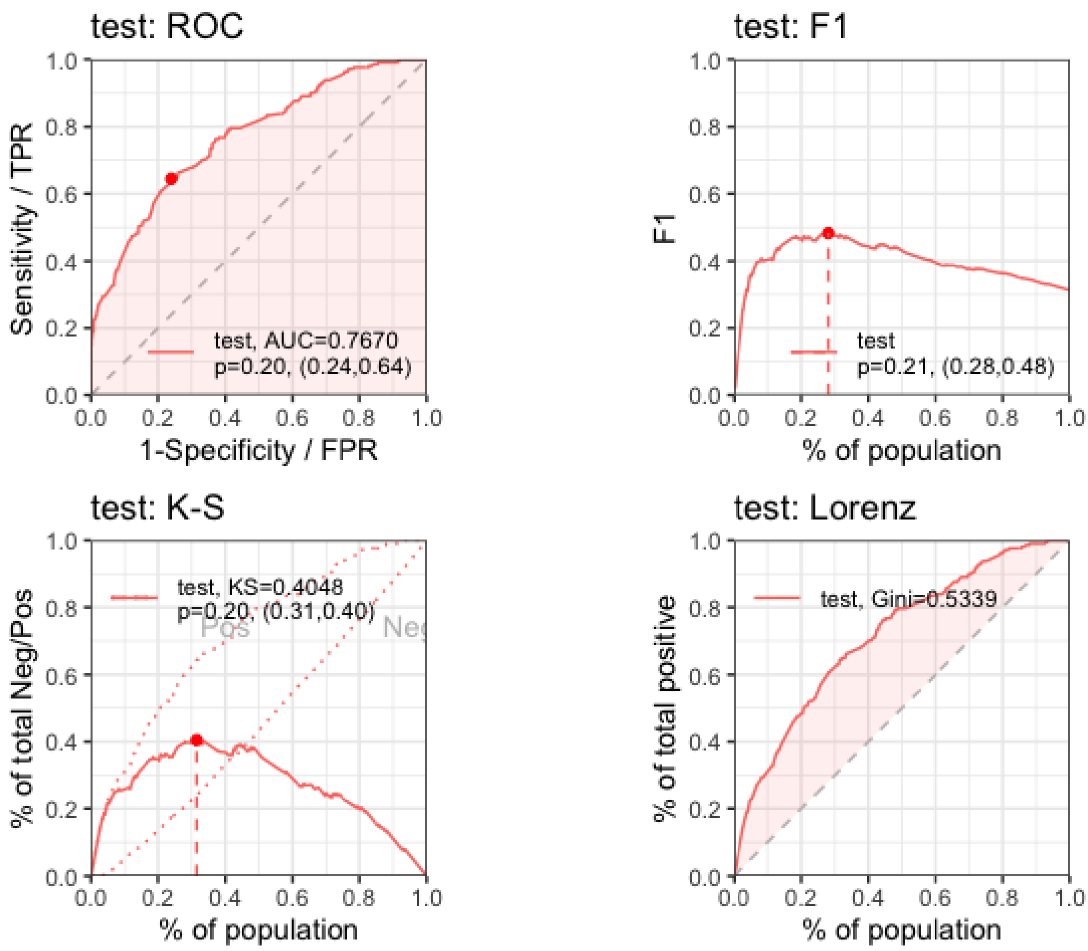

Figure 69: Model Results

Using the threshold of the first method, we obtain the following confusion matrix in Table 10.

		Predicted	
		0	1
Actual	0	694	232
	1	74	137

Table 10: Confusion Matrix

We also check for multicollinearity using the function `scorecard::vif()` to obtain the Variance Inflation Factor. The variance inflation factor again is computed as

$$\text{VIF}_k = \frac{1}{1 - \widehat{R_k}^2},$$

where $\widehat{R_k}$ is the adjusted R-squared for the regression of X_k on the other covariates. Table 11 in the following page shows the variance inflation factor for the variables. A score close 1 indicates no multicollinearity while a score above 5 indicated the presence of high levels of multicollinearity in the model. Since all values are near 1, we can conclude that there is no multicollinearity in the model.

Variable	VIF
Credit_Ratio_woe	1.020903
Credit_Status_woe	1.042796
Dependents_woe	1.025849
Education_woe	1.008359
Gross_Salary_woe	1.377545
House_Type_woe	1.034075
Months_Loan_woe	1.442858
Purpose_Loan_woe	1.018565
Yrs_Employed_woe	1.038172

Table 11: Variance Inflation Factor

6.4 Scorecard Development

To convert our logistic model into a scorecard, we will again use the following formula

$$\begin{aligned}\text{Score}_j &= \text{Offset} + \text{Factor} \times \ln\left(\frac{1-\pi_j}{\pi_j}\right)\\ &= \underbrace{[\text{Offset} - \text{Factor} \times \beta_0]}_{\text{Base Score}} + \sum_{m=1}^{k} \underbrace{-\text{Factor} \times \beta_m \times \text{WOE}_{m,j}}_{\text{Points}_{m,j}}.\end{aligned}$$

Given two points (odds_s, s) and $(2 \times \text{odds}_s, s + p_2)$, we can solve for the factor and offset as follows

$$\begin{aligned}\text{Factor} &= \frac{p_2}{\ln 2}\\ \text{Offset} &= s - \text{Factor} \times \ln(\text{odds}_s).\end{aligned}$$

Similar to the first scorecard, we use $s = 600$, $\text{odds}_s = 50$, and $p_2 = 20$. Using these, we get

$$\begin{aligned}\text{Factor} &= 28.8539\\ \text{Offset} &= 487.1229\\ \text{Cutoff} &= 529.1591\end{aligned}$$

From these values, we obtain the final credit scorecard model in Table 12 in the following page.

Variable	Bin	Points
Credit Ratio	$(-\infty, 0.06)$	36
	$[0.06, 0.08)$	10
	$[0.08, 0.26)$	0
	$[0.26, -\infty)$	-31
Credit Status	Current	3
	Non-earning	-140
	Paid-off	86
Dependents	$(-\infty, 1)$	10
	$[1, 5)$	-7
	$[5, \infty)$	23
Education	College, Vocation	3
	Elementary	-134
	High school	-45
Gross Salary	$(-\infty, 40000)$	-2
	$[40000, 47000)$	2
	$[47000, 70000)$	5
	$[70000, \infty)$	8
House Type	Missing Info	1
	Owned	12
	Relatives or Rented	-8
Months Loan	$(-\infty, 17)$	41
	$[17, 21)$	16
	$[21, 24)$	7
	$[24, \infty)$	-6
Purpose Loan	Appliance	-8
	Auto	-6
	Leisure or Travel	1
Years Employed	$(-\infty, 5)$	11
	$[5, 8)$	-1
	$[8, 11)$	-14
	$[11, 20)$	-10
	$[20, \infty)$	9
Base Points		530

Table 12: Credit Scorecard - Non-Essential

References

[1] Siddiqi, Naeem. Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards. Second edition, Wiley, 2017.

[2] Ternary Search - The Computer Science Handbook. https://www.thecshandbook.com/Ternary_Search. Accessed 16 Nov. 2021.