



# Career that Tend to be Unpaid for Motorcycles Sales Loans

Mari Atsuki<sup>(✉)</sup> and Yumi Asahi

Department of Management System Engineering,  
School of Information and Telecommunication Engineering,  
Tokai University, Tokyo, Japan  
5bjml120@mail.u-tokai.ac.jp, asahi@tsc.u-tokai.ac.jp

**Abstract.** Country A is in Latin America and GDP is lower than average. Prices of Japanese products such as motorcycles are on an upward trend. It was influenced by instability of the world economy. The customer selects loan payment. Among them, some customers cannot pay for the specified payment period. If this situation deteriorates, it will be difficult to recover manufacturing costs. Therefore, we analyze the characteristics of customers who loan bankruptcy.

There is a weak positive correlation (0.300) between average income for each province and sales volume. There is a negative correlation (-0.542) to the main income amount and those who go bankrupt. So it can be said that people living in poor countries cannot easily buy a motorbike. Even if they make loans to purchase, they cannot pay. I divided the state data into five. It turned out that there was economic disparity. The southern region is rich and the north region is a poor region. There are many customers who cannot repay the loan to the north region.

In this country, 13% of 10 million young people have not enrolled in school and are not working. In order to see the relationship between academic background and other factors, we quantified the academic qualification. There is a negative correlation between academic records and the proportion of unpaid -0.673. Therefore, I can say that the lower the academic background, the more delayed repayment of the loan. I conducted a multiple regression analysis with the loan's unrepayable party as the dependent variable. The standard partial regression coefficients were -0.543 for educational record number and -0.327 for main income. And I learned that educational background is a factor that has a big influence on judging whether it will be unpaid than main income.

We analyze logistic regression on the probability that that person will be Bad with these elements. We understood what condition customers are likely to be Bad. My goal is to be able to know the percentage of probabilities that your loan can not be repaid from your profile information.

**Keywords:** Career · Loan · Loan bankruptcy · Educational history  
Main revenue · State

## 1 Introduction

World finances were influenced by Lehman shock in 2008. In the spring of 2009, the economy had bottomed out, and, it had a gradual recovery trend. In August 2011, financial instability was increased by downgrade of US government bonds and impact

of the deepening of the European debt crisis. In Latin America, major currencies plunged. And the price of Japanese products in Latin America has been on an upward trend. Hence customers choose installment payment.

Among them, some customers can't pay in the specified payment period. The country's poverty alleviation is the cause. The people's income has increased. And people who were able to afford to some extent purchased loans and purchased products. They don't have loan experience and understanding of contract contents is often inadequate. Therefore, the consciousness to the loan is low, leading to bankruptcy. In addition, some people make new loans even though loan repayment has not been paid. They still make loans even if the down payment gets higher. This is an international issue. If this situation deteriorates, manufacturers have difficulty in recovering manufacturing costs. Thus we analyze the characteristics of customers who will be outstanding.

In this study, we use motorcycles sales data from country A. GDP of country A is lower than average. Motorcycles is used as a practical means of transportation in country A. And it is essential for economic development such as commuting and agriculture.

## 2 Data Summary

We use motorcycles sales data from country A in September 2010 to June 2012. There are 14,304 data and 359 variables. For example, borrowing amount, birthday, resident state, revenue, career and whether the loan has been repaid or not.

There is not has data of age, and we calculated it using data of application date and birthday. There are some blanks in the main income data. Then supplementing that data with a rating score without missing values. The way is like this. Collect other data with the same value as the rating score of the data whose main income is blank. And we calculate the average value of those main incomes. We put it in the main income amount that was blank.

And we cleaned up the data. We deleted data with blank spaces in key variables such as interest amount and borrowings. With this, the number of data become 13,214.

From now on we will call people who failed to repay the loan as "Bad". People who have once become bad continue to be bad. Therefore, the proportion of data is increasing. Of all the data, 6 months Bad is 9.1%, 12 months Bad is 16.8%, 18 months Bad is 23.0%. As the number of data increases, it leads to more accurate analysis, thus we decided to focus on 18 months Bad. In addition, because it includes "6 months Bad" and "12 months Bad", comprehensive measures can be considered.

There are 4 state variables of state (resident state, resident registration state, working state and first trust survey state.) The same data for all four registered states is 81.8%. The same data for three registered states is 17.1%. Among them, only the resident state is rarely registered in a different state. Therefore, resident state data is the most stable. We use the resident state as state data on that person data.

We made a group of main income. We did it based on the income stratum rank index issued by the economic foundation of country A. The average primary income for Country A is 3,010.48. The A/B layer is wealthy class, the average main income is

145,594.12. The C layer is the middle class, the average main revenue is 2,802.01. The D and E layer are poor and the average of the main income is 1,323.34 and 836.46 respectively. You can see that there is economic disparity.

### 3 Fundamental Statistics

According to the study of Miyoshi Yusuke (2013), regional features are involved. For example, regional wage rate, security, percentage of young people. Thus we see at the data by state.

#### 3.1 Main Revenue

There is a weak positive correlation (0.300) between the average amount of major revenue by state and sales volume. There is a negative correlation (-0.542) by Bad people. Therefore we can say that people living in poor states can't easily buy motorcycles. Even if they make a loan to purchase, they will not be able to pay.

We classified state data as into division of 5, the North, the Northeast, Midwest, the South and Southeastern area based on a geographic division of the Ministry of Agriculture, Forestry and Fisheries. We corresponded it to the state data arranged in order of the main income. It is shown in Fig. 1. We understand the following. There is an economic disparity that the poor increase from the South to the North. Southeast sales are particularly high. The number of units sold varies by state. The sales volume of North-A is small, however the proportion of Bad is very high.

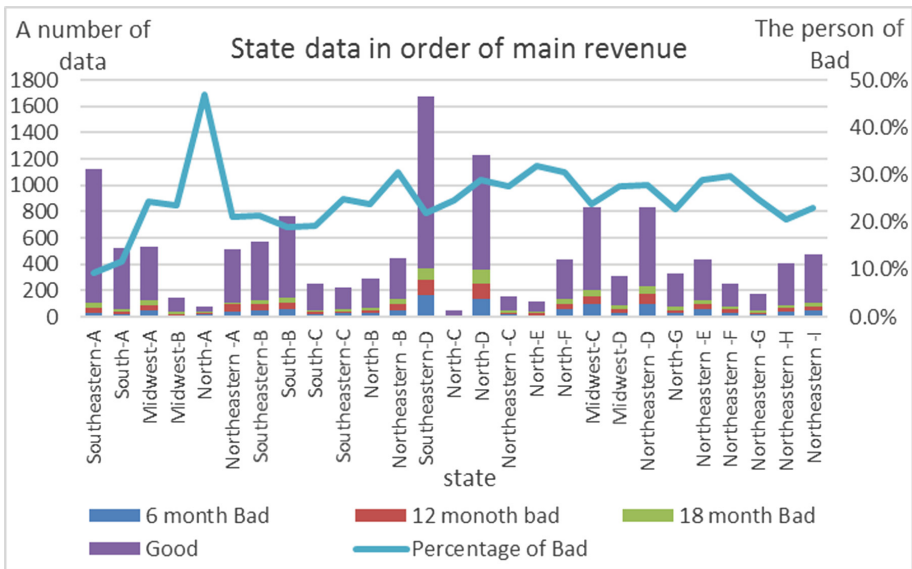


Fig. 1. State data in order of main revenue

We compared the highest region with the lowest region in average of major revenue. The region with the highest main income is the southeastern part, which is 1784.807. The region with the lowest main income is the north part, which is 1483.431. As expected, there are many bad customers in the north. They are 18.3% of the southeastern population, 28.5% in the north population. We think that the cause of the difference is in occupation. Self-employed is a job with little income and stability. There are more self-employed in the northeastern part than in the southeastern part. Hence even though the northern part has fewer borrowings than the southeastern part, there are many customers who cannot repay the loan.

### 3.2 Educational History

Another cause of the economic disparity is that 13% of 10 million young people are not enrolled in school, and they are not working yet. In this data of this study, 2.3% is “data without educational history”. Considering that people who cannot afford to receive education can think that they cannot afford to purchase motorcycles, it is a convincing data.

In country A, ages 6 to 11 go to Basic school, then until 12 years old is Secondary school, then until 18 years old is High Education. To see the relationship between educational background to other elements, we quantified my academic background. We did it in the following way. Number each academic record in turn. (Basic school: 1, Secondary education: 2, Higher education: 3, Correspondence training: 4). In case of dropout, subtract 0.5 from each number. There is a negative correlation of  $-0.651$  between the academic record and the ratio of unpaid. Therefore we can say that the

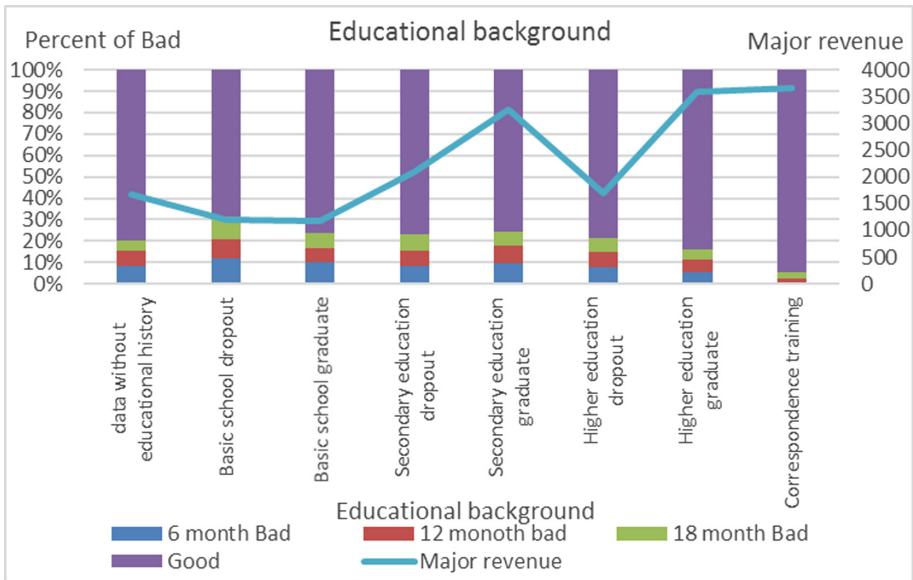


Fig. 2. Educational background data

lower the academic background, the more the repayment of the loan is delayed. However, the correlation between educational background and main revenue is 0.402. The reason is probably the difference between dropout and graduation. The main income amount is higher for secondary education graduation than for higher education dropout. Even if you have a high academic background, you probably cannot get a good job if they dropout. In addition, primary income of people without educational history is higher than those who were in the basic school and the secondary education dropout. This is because someone who has not attended school has been working for a long time. It is because there are advantages in terms of time (Fig. 2).

## 4 Multiple Regression Analysis

We conducted multiple regression analysis with the proportion of Bad persons by state as the dependent variable. Independent variables are the average of educational background and main income. Because of multiple regression analysis, the weight determination coefficient was 0.542, which was a significant value at 1% level. The standard partial regression coefficients were  $-0.543$  for educational record number and  $-0.327$  for main income. And I learned that educational background is a factor that has a big influence on judging whether it will be unpaid than main income.

## 5 Logistic Regression Analysis

We used logistic regression. The difference between logistic analysis and multiple regression analysis is the data type of the objective variable. The objective variable of multiple regression analysis is numeric. Logistic regression analysis is nominal scale data. In the multiple regression analysis described above, “how much percentage of people in the state would be Bad” was found. Logistic regression is seen more concretely than that. We know the probability that the person himself will be Bad. Therefore, it meets the purpose of this research. The judgment criterion of this analysis result is that the significance probability is less than 0.05. When the odds ratio (Exp (B)) is larger than 1, it can be said that the larger the explanatory variable, the more it affects the target variable. Conversely, when the odds ratio (Exp (B)) is smaller than 1, it can be said that the larger the explanatory variable, the less influence the target variable.

The objective variable is Bad. The dependent variable is a career. We analyze by main revenue, education history, resident state in order. Main income data has low regression coefficient. In addition, the main income level is four stages, so you can narrow down the target roughly. There are 8 levels of educational history and 26 states. Therefore we can see in detail in order.

### 5.1 Main Revenue

First, the main revenue analysis the dependent variable. Logistic regression requires that the dependent variable be a binary variable. We decided to use the data was

grouped main revenue level. We made dummy variables of main income level. These are dependent variables. However, the data of the AB layer is 1.11% (147 data) of the total. Furthermore, only 19 data are bad in the AB layer. We excluded it because the number of data is too small.

**Table 1.** Analysis result of main revenue dependent variable

	B	Standard error	Wald	df	Significance probability	Exp(B)
E layer	0.651	0.248	6.884	1.000	0.009	1.918
D layer	0.935	0.248	14.258	1.000	0.000	2.548
C layer	0.209	0.252	0.686	1.000	0.407	1.232
constant	1.908	0.246	60.203	1.000	0.000	0.148

The results are shown in Table 1. Among them, the significance probability is 0.05 or less only in layers D and E, or the odds of these two layers can be said to be reliable. Odds exceed 1 in all layers. In other words, it can be said that it tends to be Bad if it is this layer. The odds of the D layer is higher than that of the E layer. In other words, it can be said that E layer is less likely to be Bad than D layer. The reason is that motorcycle makers are taking measures against E customers. Therefore, among the poor, the D layer with more income than the E layer tends to be Bad. The C layer has a high significance probability and odds are also relatively low. To see more about Bad, we focus on the D and E layers and proceed with the analysis.

### 5.2 Educational History

We made dummy variables for each academic background too. These are dependent variables. The data of correspondence education is 0.36% (38 data) of all data. Only 4 cases are Bad in that. So exclude this variable.

The results are shown in Table 2. The significance probability was less than 0.05 and the significant variables were those related to “Basic School” and “Secondary School”. All odds are over 1. The lower the academic background is, the more it tends to be Bad. We find that the lower the educational background is, the more likely it is Bad. In each degree, it turns out that the person who is dropping out tends to be Bad more than the person who is Graduate.

The chapter of fundamental statistics shown that the main income of those who do not have an educational history is high. As can be inferred from there, the odds of data without educational history were lower than basic school education and secondary school education. It is a special case that there is no history of education. In addition, the odds of dropped people are higher than those who graduated. It turned out that those who dropped out from this result were likely to be Bad.

**Table 2.** Analysis result of educational history dependent variable

	B	Stand- ard error	Wald	df	Sig- nificance probability	Exp(B)
No edu- cational his- tory	0.714	0.462	2.391	1.000	0.122	2.042
Basic school dropout	1.323	0.473	7.829	1.000	0.005	3.755
Basic school graduate	0.958	0.446	4.613	1.000	0.032	2.606
Second- ary education dropout	0.969	0.444	4.765	1.000	0.029	2.636
Second- ary education graduate	1.048	0.433	5.846	1.000	0.016	2.851
Higher education dropout	0.829	0.444	3.491	1.000	0.062	2.291
Higher education graduate	0.644	0.443	2.110	1.000	0.146	1.903
constant	- 2.100	0.433	23.575	1.000	0.000	0.122

As stated above, the enrollment rate in this country is low. Up to the second degree is compulsory education. However, the school truancy rate and voluntary dropout rate are still high. The reason is often in the environment of parents and family. It becomes impossible to attend school with unstable work and accompanying move. Different family problems such as divorce and divorce. In addition to these, parents are also less aware of their education. We think there is poverty in these backgrounds. They cannot get a stable job and cannot prepare the environment. Living expenses are given priority, there is no educational expenses. For these reasons, it is possible to assume that a person who dropped out is a poor household whose parents' educational awareness is low and who is not in a stable position. In addition, educational institution infrastructure is sometimes inadequate. You can see that there is no money in the country or state. Because it is a poor area it is difficult to get a stable job. And it's difficult to solve educational problems more.

Anyway, we can see that basic school and secondary education are factors that tend to be Bad. We focus on these and proceed with the analysis.

### 5.3 State

Finally, we analyzed the state data as an explanatory variable. We made a dummy variable like educational background.

The results are shown in Table 3. There were 11 states with significant probability less than 0.05.

**Table 3.** Analysis result of state dependent variable

	B	Stand- ard error	Wald	df	Sig- nificance probability	Exp(B)
South- eastern-A	- 1.301	0.187	48.579	1.000	0.000	0.272
South- A	- 0.944	0.194	23.593	1.000	0.000	0.389
Mid- west-A	- 0.230	0.161	2.044	1.000	0.153	0.795
Mid- west-B	- 0.338	0.309	1.191	1.000	0.275	0.713
North- A	- 1.287	0.338	14.503	1.000	0.000	3.623
North- eastern -A	- 0.400	0.169	5.629	1.000	0.018	0.670
South- eastern-B	- 0.392	0.179	4.822	1.000	0.028	0.676
South- B	- 0.440	0.156	7.965	1.000	0.005	0.644
South- C	- 0.463	0.219	4.462	1.000	0.035	0.630
South- eastern-C	- 0.068	0.213	0.102	1.000	0.750	0.934
North- B	- 0.244	0.191	1.624	1.000	0.203	0.784
North- eastern -B	- 0.008	0.160	0.003	1.000	0.959	0.992
South- eastern-D	- 0.273	0.135	4.103	1.000	0.043	0.761
North- C	- 0.137	0.410	0.111	1.000	0.739	0.872
North- D	- 0.023	0.137	0.027	1.000	0.869	0.978
North- eastern -C	- 0.041	0.235	0.031	1.000	0.860	0.959
North- E	- 0.452	0.261	2.986	1.000	0.084	1.571
Mid- west-C	- 0.329	0.148	4.930	1.000	0.026	0.719
Mid- west-D	- 0.130	0.186	0.485	1.000	0.486	0.878
North- eastern -D	- 0.014	0.142	0.010	1.000	0.921	0.986
North- G	- 0.426	0.203	4.381	1.000	0.036	0.653
North- eastern -E	- 0.058	0.165	0.124	1.000	0.725	0.944
North- eastern -F	- 0.051	0.187	0.076	1.000	0.783	1.053
North- eastern -G	- 0.288	0.225	1.641	1.000	0.200	0.750
North- eastern -H	- 0.409	0.176	5.386	1.000	0.020	0.664
North- eastern -I	- 0.371	0.173	4.576	1.000	0.032	0.690
con- stant	- 0.802	0.115	48.365	1.000	0.000	0.449



Among them, only North-A had an odds of 1 or more. The proportion of Bad in North-A is the highest at 46.9%. The overall Bad rate is 23.1%. But the average primary income is fifth highest. Examining the data, there were 2 data on 5-digit income. We understand this is the cause. The average value without these two data was 1662.19. This is a lower value than the average and it can be said to be a poor state.

The significance probability was 0.05 or less, and the odds other than North-A were all less than 1. That is, other significant variables can be said to be “It is hard to become Bad if you live in that state.”

This state is located at the northernmost point of country A. It is a region of climate with high temperature and humidity which is directly under the equator. The average temperature is over 27 °C and annual rainfall is 3500 mm. Many natural remains. Many people sell their timber to make a living. However, it causes the flooding of the river. It will be a further burden to get the damage. Therefore, it becomes Bad in a place where households are hard to stabilize.

The lowest odds is Southeastern-A. In other words, it can be said that it is a state that is least likely to be Bad. Actually the proportion of Bad is 9.4%. It is clear that the economy is stable because the average value of main income is remarkably high as 15530.76.

The reason why economic conditions are good is that it is the center of industry, commerce and finance in country A. It is blessed with climate, infrastructure such as roads, railroads and harbors are in place. The industry has developed and has a large population. Therefore, there are many commercial jobs. There are many wealthy people who have stable jobs. So they are easy to pay off when they form a loan.

The average educational number is 17.04 for State North-A and 21.55 for Southeastern-A. In other words, people in North-A are more likely to drop out than to graduate from secondary degree. Many people in Southeastern-A graduate from secondary degree. There are also people who go on to further advance.

We understand again that it relates to educational history and Bad and main income. And there is also an influence of the economic state of society. For the society to develop, it may be affected by weather and others.

## **6 Discussion/Summary**

We analyzed what kind of career affect to become Bad. Therefore level of revenue is D/E level, academic background is graduation of secondary educational background from dropout of basic school and those living in North-A are prone to Bad. If a customer who meets these conditions wishes to form a loan, should be wary. If we do decision tree analysis, it will be easier to understand. We will make this a future subject. In that case, if you also put variables on other backgrounds, accuracy will improve. Then because there is a possibility that the data is insufficient we have to perform statistics.

## References

- Ministry of Economy, Trade and Industry: Trade White Paper 2012 Chapter 1 World Economic Trends Section 6 Latin America, Russia Economy (2012)
- Miyoshi, Y.: Empirical study of inverse selection in consumer finance market. Kagawa Univ. Econ. Rev. **88**, 537–542 (2016)
- Ministry of Agriculture, Forestry and Fisheries: Heisei 21 year Country Report. <http://www.aff.go.jp/primaff/koho/seika/project/pdf/nikokukan12-3.pdf>. Accessed 18 Dec 2017
- Ministry of Education, Culture, Sports, Science and Technology (MEXT): Survey on reforms on academic systems in various countries in FY2005 (2). [http://www.mext.go.jp/a\\_menu/shougai/chousa/1351481.htm](http://www.mext.go.jp/a_menu/shougai/chousa/1351481.htm). Accessed 17 Dec 2017