

QnA Analysis of the Drivers of Loans defaults

What are the critical drivers of loan defaults

John Karuitha

Friday April 21, 2023

RQ1 Does default rate depend on the purpose of the loan?

SQ 1.1 Give the variables represented in each role in Figure 1

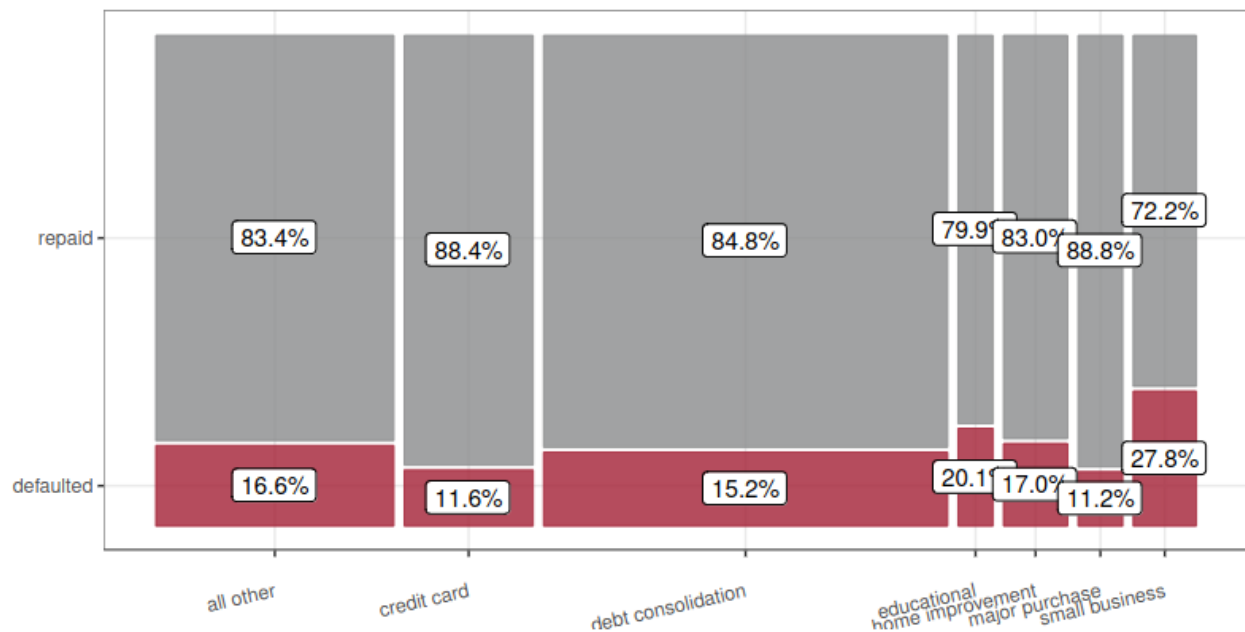


Figure 1: Default rate as a function of the loan's purpose

- Explanatory: Purpose
- Response: Defaulted

SQ 1.2 Fill in the blanks and circle the correct options in parentheses to complete the text summary

Default rate differs somewhat by loan purpose. The highest default rate was for **debt consolidation** loans: 16.764 % of them defaulted. The lowest default rate was for **major purchase** loans: 88.787 % of them were repaid in full. Perhaps surprisingly, the default rate for credit card loans was 8.5 percentage points (higher / lower) than the default rate for educational loans. In other words, educational loans defaulted at a rate 8.5 % (higher / lower) than credit card loans.

RQ 2 Does the debt-to-income ratio seem to influence the interest rate borrowers receive?

SQ 2.1 For this research question, what are

a. The variables?

The variables are debt to income (dti) and interest rates (int_rate).

b. Variable types?

Both debt to income (dti) and interest rates (int_rate) are quantitative /numeric variables.

c. Variable roles?

The interest rate (int_rate) is the response (dependent) variable. The debt to income (dti) is the explanatory (independent) variable.

d. Appropriate type of chart?

Because there are 2 numeric variables, the appropriate chart is the scatter plot, with interest rate on the y-axis and debt to income ratio on the x-axis.

e. Appropriate summary statistic(s)?

The correlation coefficient and covariance would be useful summary statistics to quantify the degree of linear relationship between interest rates and debt to income ratio.

SQ 2.2 Chart

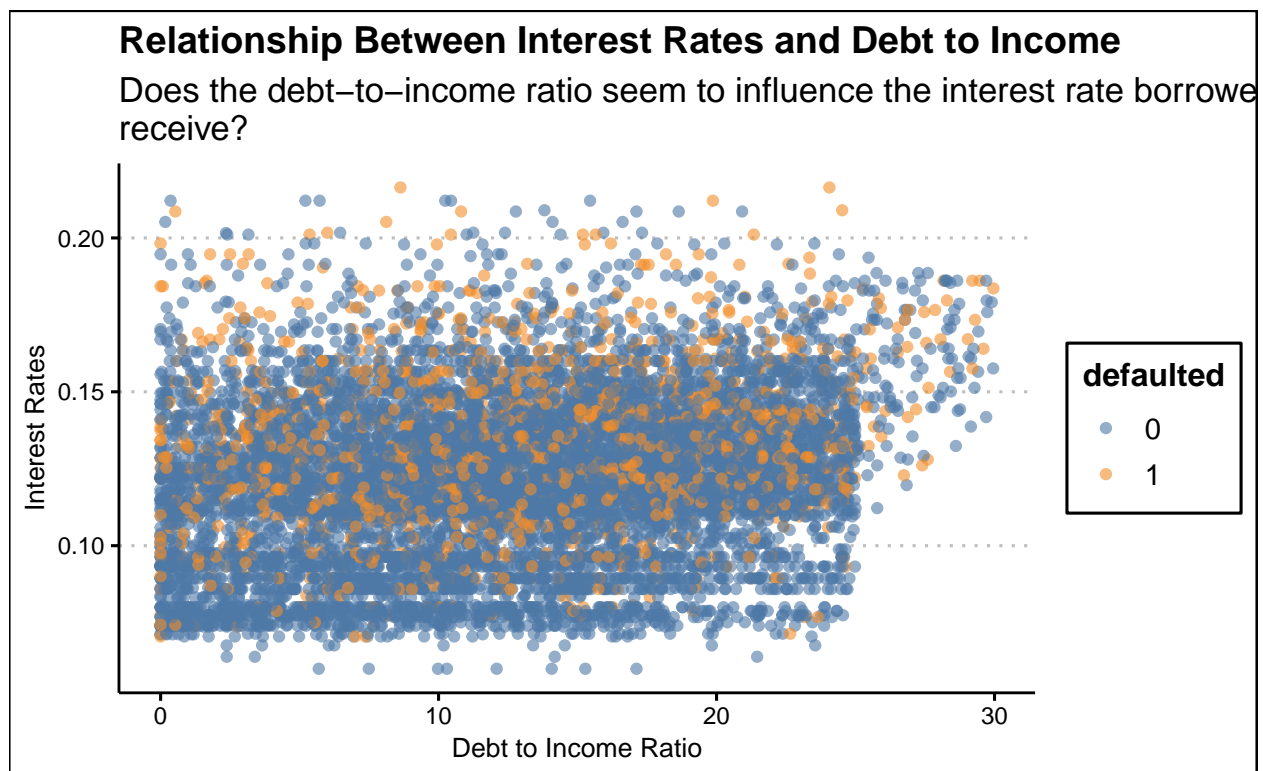


Figure 2: Relationship Between Interest Rates and Debt to Income

SQ 2.3 Text summary—give a qualitative summary plus one summary statistic

Qualitative summary: There is little relationship between interest rates and debt to income ratio. Note that the scatter plot is approximately flat meaning that as debt to income ratio goes up, interest rates on loans do not change significantly.

Quantitative summary: The correlation coefficient between interest rates and debt to income ratio is 0.22.

RQ 3 How is interest rate related to FICO score?

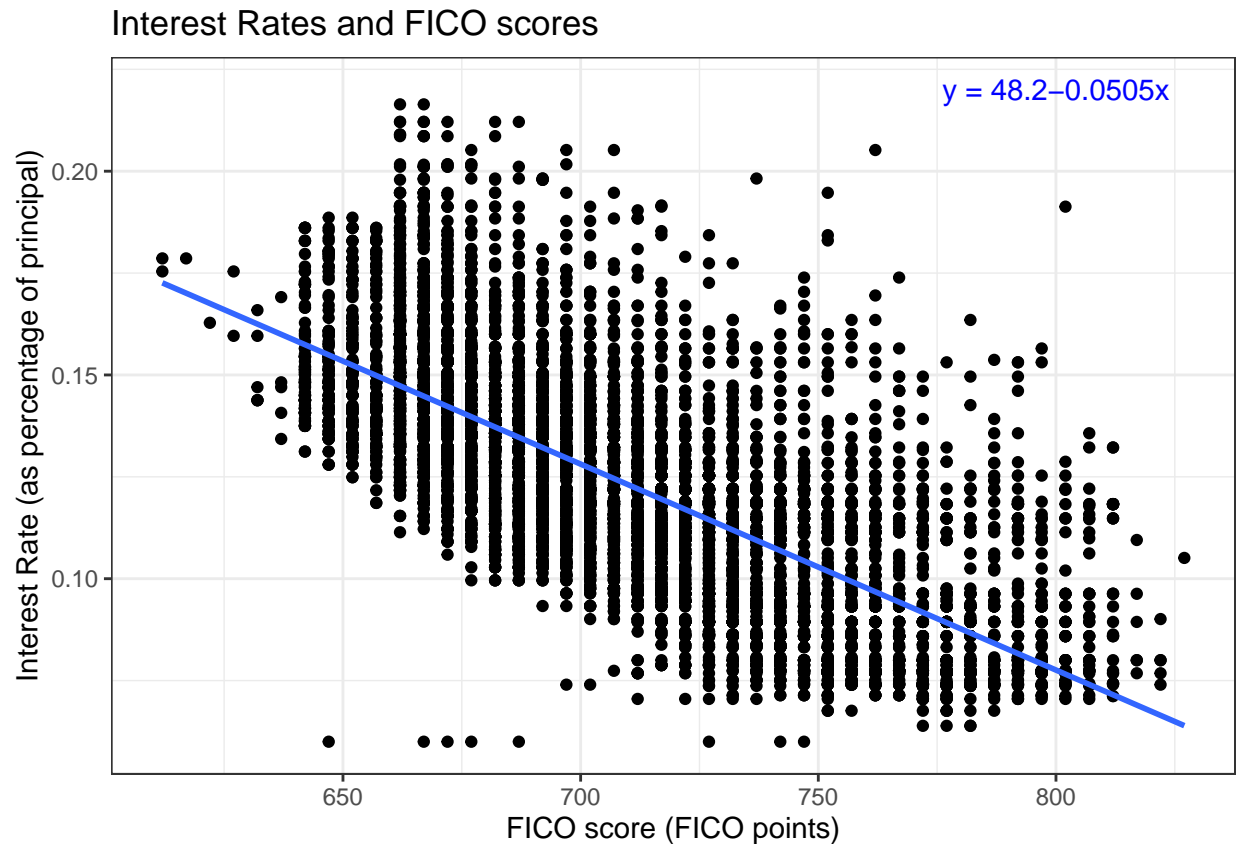


Figure 3: Chart visualizing the relationship between FICO score and interest rate on the loan. A linear model is shown in blue, together with its formula.

SQ 3.1 Text summary—with qualitative and quantitative summaries (use the slope of the line)

Qualitative summary: There is a notable negative correlation between FICO scores and interest rates. Specifically, when the FICO score rises by a percentage point, interest rates correspondingly go down by an average of 5.05 percentage points. This is the reason the plot has a relatively steep downwards slope.

Quantitative Summary: We summarise this relationship using the correlation coefficient as there are two quantitative variables. The correlation coefficient is -0.715, indicating a relatively strong negative relationship between FICO scores and interest rates.

RQ 4 Can the borrower's annual income be used to predict a loan's monthly instalment amount?

SQ 4.1 For this research question, what are

a. The variables?

The variables annual income (annual_inc) and monthly instalment amount (instalment).

b. Variable types?

Annual income is a quantitative variable.

Loan instalment is also a quantitative variable.

c. Variable roles?

In this context, monthly instalment amount is the response variable. Annual income is the explanatory or independent variable.

d. Appropriate type of chart?

Because these two variables are quantitative, the scatter plot is the most appropriate chart to visualise the data.

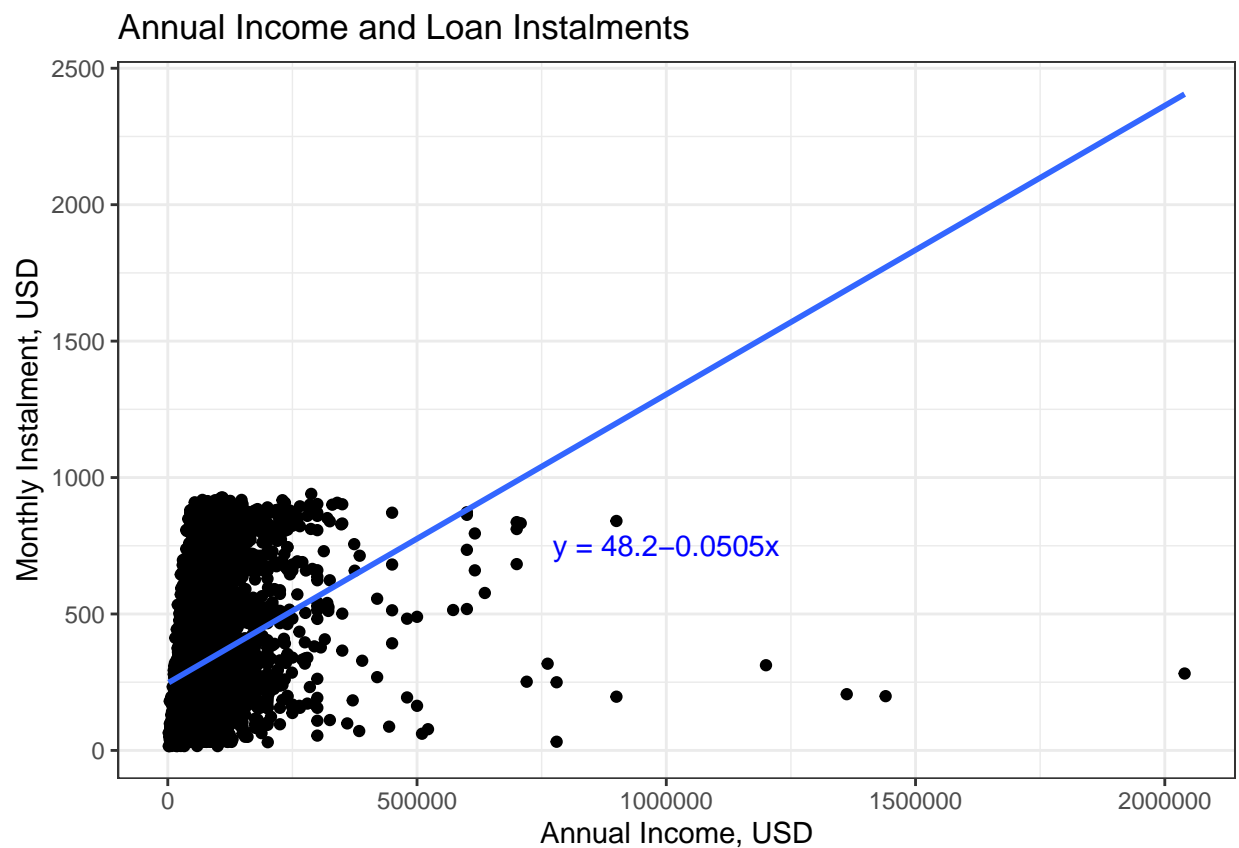


Figure 4: Annual Income and Loan Instalments

e. Appropriate summary statistic(s)?

These two are quantitative variables and hence, I use the correlation coefficient. The correlation between annual income and monthly instalments is 0.313.

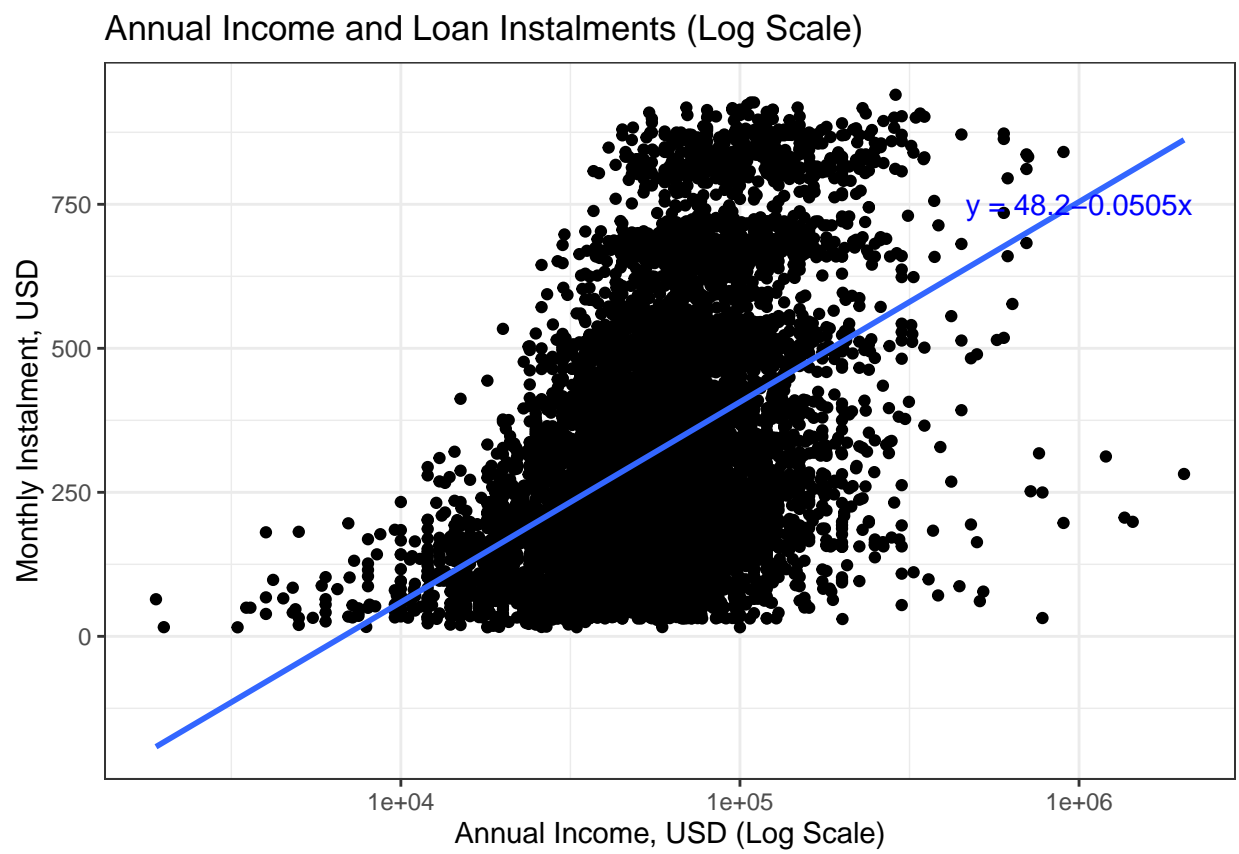


Figure 5: Annual Income and Loan Instalments (Log Scale)

In addition, I run regression analysis (Table 1). Here, I use both the logarithm of income and the raw income variables.

Table 1: Regression Analysis

	<i>Dependent variable:</i>	
	installment	
	Raw Model (1)	Log Model (2)
annual_inc	0.001*** (0.00003)	
log(annual_inc)		151.000*** (3.080)
Constant	247.000*** (3.010)	-1,331.000*** (33.700)
Observations	9,578	9,578
R ²	0.098	0.201
Adjusted R ²	0.098	0.201
Residual Std. Error (df = 9576)	197.000	185.000
F Statistic (df = 1; 9576)	1,039.000***	2,406.000***
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

To answer this research question, you could use either annual income or the natural log of annual income as the predictor. Assume you would like to use a linear model.

SQ 4.2 Which variable would make a better predictor and why?

Circle the best answer a. or b. and explain in the box below.

a. Annual income.

b. Natural log of annual income and why?

The model with the natural logarithm of income as the outcome has better predictive power. looking at the adjusted R-squared, the variability in the logarithm of annual income can explain 20.1% of the variability in monthly instalments compared to 9.8 % explained by the model using raw annual income.

RQ 5 Does the number of times the borrower has been 30+ days late provide an indication of whether they will default?

*SQ 5.1 For this research question, what are

- a. The variables?

The variables are:

- Defaulted. Did the borrower default?
 - delinq_2yrs: The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
- b. Variable types?
- Defaulted is a categorical variable with 2 levels 0 and 1.
 - delinq_2yrs: This is a quantitative discrete variable of the count of times the borrower has been 30+ days past due on a payment in the past 2 years.
- c. Variable roles?
- Defaulted is the response or dependent variable.
 - delinq_2yrs is the explanatory or independent variable.
- d. Appropriate type of chart?

Here, we are dealing with a quantitative and categorical variable. We can use a box plot to visualize the incidence of delinquency between defaulters and non-defaulters.

- e. Appropriate summary statistic(s)?

In this case, I would compute the mean and median number of times that defaulters and non-defaulters were 30+ days past due on a payment in the past 2 years.

SQ 5.2 Chart

See Figure 6 below.

SQ 5.3 Table

Table 2: Number of Loan Delinquencies, Defaulters vs Non-defaulters

defaulted	Mean_times	Median_times
Non-defaulters	0.162	0
Defaulters	0.175	0

Note:

Non-defaulters have marginally low incidence of average delinquency

SQ 5.4 Text Summary

The difference in the average delinquency rate between defaulters and non-defaulters is marginal. For instance, the median number of delinquency between the two groups is zero. However, non-defaulters show a marginally lower average number of delinquencies in the past 2 years, at 0.162 times and 0.175, respectively.

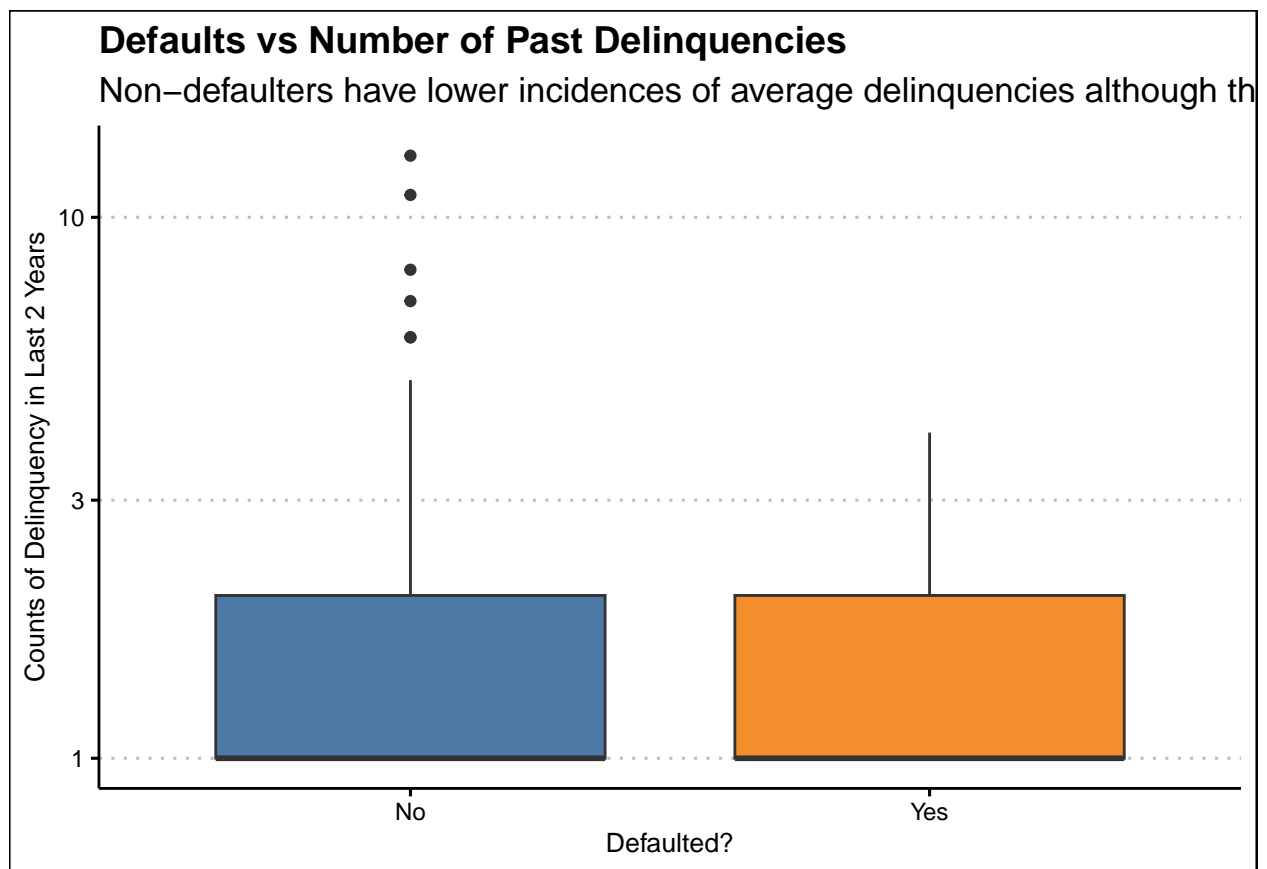


Figure 6: Defaults vs Number of Past Delinquencies

RQ 6 Does the purpose of a loan seem to affect the interest rate?

SQ 6.1 For this research question, what are:

a. The variables?

The variables are:

- Purpose (purpose).
- Interest rates (int_rate).

b. Variable types?

- Purpose (purpose) is a categorical variable with 7 levels or categories.
- Interest rate (int_rate) is a quantitative, continuous variable.

c. Variable roles?

- Interest rate is the response or dependent variable.
- Purpose is the explanatory or independent variable.

d. Appropriate type of chart?

Again, given that we have a categorical and continuous variable, a box plot is suited for visualization.

e. Appropriate summary statistic(s)?

In this case, it would be appropriate to compute the mean and median interest rates per loan purpose.

SQ 6.2 Table

Table 3: Interest Rates by Loan Purpose

purpose	mean_interest_rates	median_interest_rates
small_business	0.138	0.138
debt_consolidation	0.127	0.128
educational	0.120	0.122
credit_card	0.120	0.119
home_improvement	0.118	0.118
all_other	0.117	0.118
major_purchase	0.114	0.116

Note:

Small businesses have the highest interest on loans. Major Purchases have the least

SQ 6.3 Text summary

The purpose of a loan has a notable relationship with interest rates. Small businesses face the highest interest rates. Major purchases are charged the lowest rates, presumably because these purchases can serve as collateral.

RQ 7 Is this data set consistent with the conventional wisdom regarding interest rates and default risk?

The conventional wisdom is that lenders charge higher interest rates for higher-risk loans, i.e. those that are more likely to default.

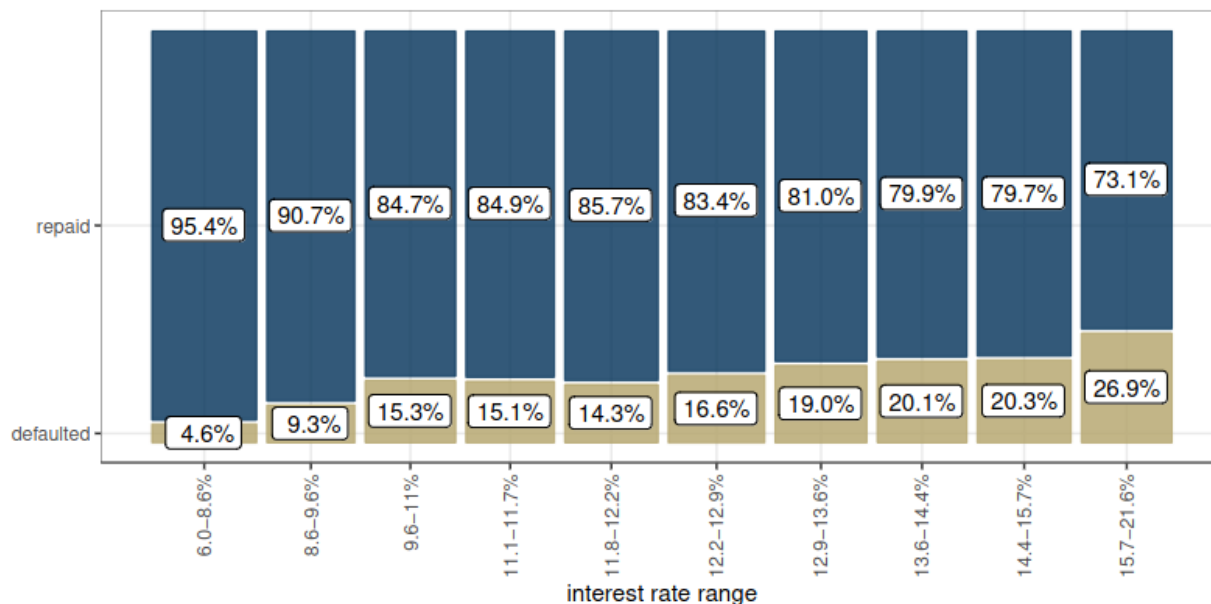


Figure 7: “Figure 3: A chart visualizing the relationship between interest rate and default.”

SQ 7.1 Based on Figure 3, is this data set consistent with the conventional wisdom?

Why or why not—as a text summary, both qualitative and quantitative, and include metadata

Qualitative: The figure is line with conventional wisdom. There is a significant relationship between interest rates charged on loans and the rate of interest. Specifically, loans with higher rates of interest exhibit a higher rate of default.

Quantitative: In the figure, we see that loans with an interest rate of between 15.7% and 21.6% have a higher default rate of 26.9%. This is in contrast with loans with an interest of between 6% and 8.6% that have a default rate of 4.6%.