Explaining Loan Default: A Data Visualization Approach

## by Lawrence Agyepong

Investigation Overview

I present the key findings from the exploratory data visualization on the factors that explain loan defaults. I investigated the relationship between loan default and the following variables: Employment Status, Income, Borrower Rate, Prosper Score, Credit Score, and Debt to Income ratio.

Dataset Overview

The objective of this project is to employ exploratory data visualization techniques to examine the factors that predict loan default. I use loan data from Prosper Marketplace containing 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others. After preliminary wrangling, the data was reduced to 42,452 rows and 10 columns.

Distribution of Loan Status

Out of the 42,452 loans, about 11% corresponding to 4,833 have defaulted and the reminder has been completed.

## Distribution of Loan Status
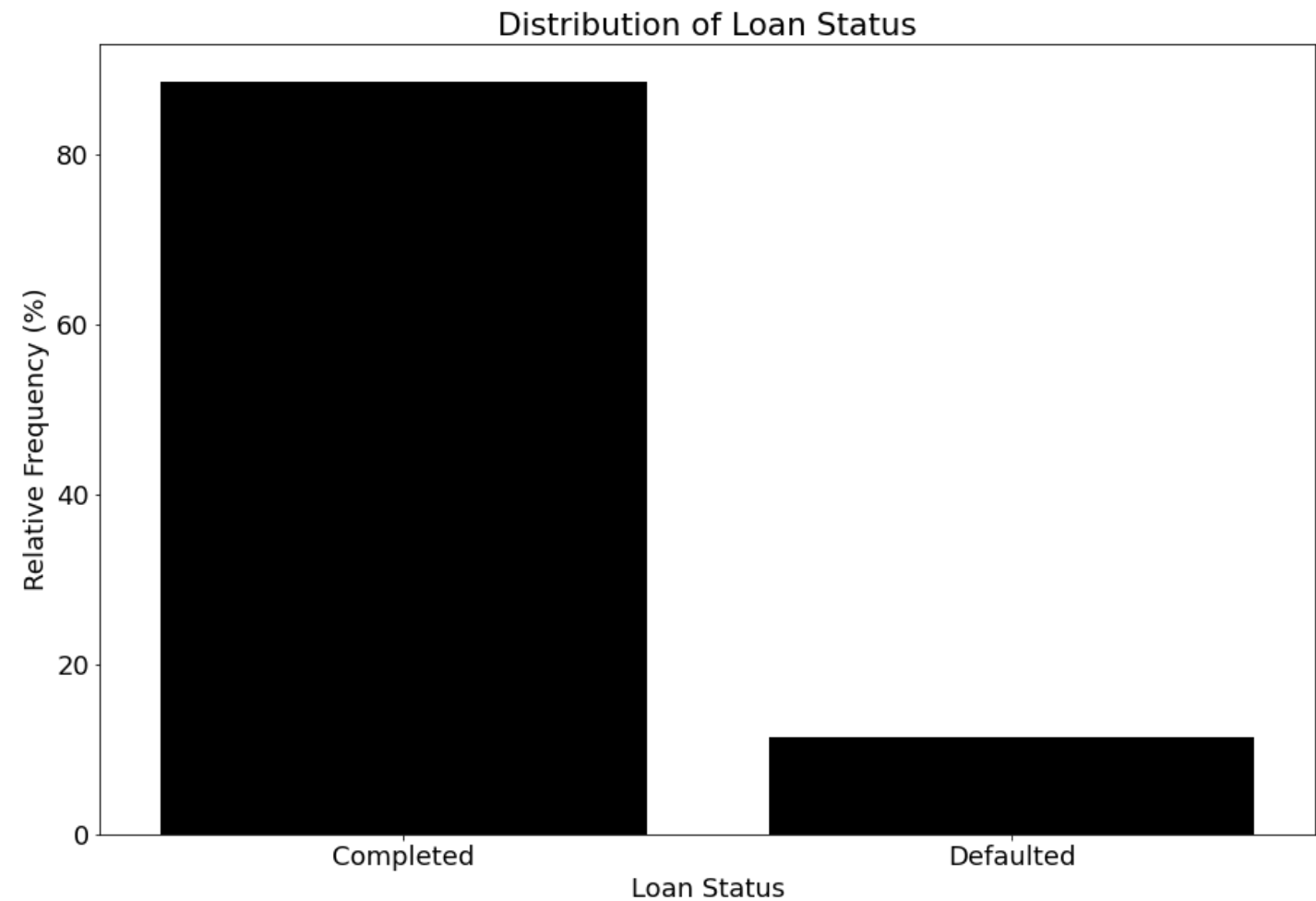
Loan Default vs. Income

Generally, we observe that higher income levels correspond to lower loan default rates. From $1-24,999 to $100,000+, loan default rate drops from 10.3% to 6.33%.

An interesting observation is that the default rate among the unemployed is lower than that of the 3 income groups between $1-24,999 and $50,000-74,999. Also, the Default rate is highest among clients whose income levels are not displayed.

Loan Default Rate across Income Levels
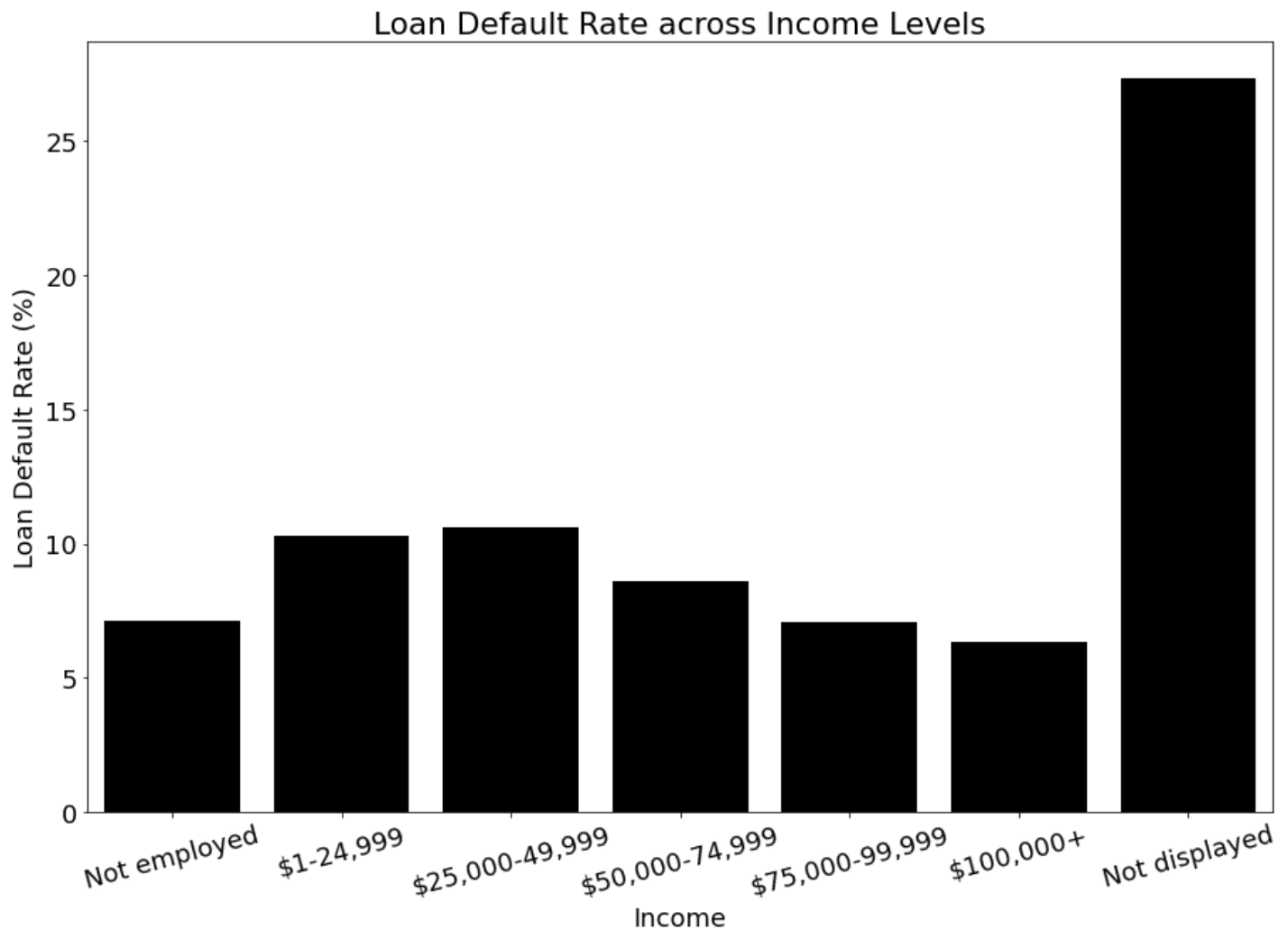
Loan Default vs. Prosper Score

There is an inverse relationship between loan default rate and Prosper Score such that, higher Prosper Scores correspond to lower levels of default rate. Except at Score 5 which has a higher default rate than its preceding level (i.e., Score 4), the default rate decreased consistently at high levels of Prosper Score. For example, the default rate dropped from about 9.8% at Score 1 to 1.25% at Score 10.
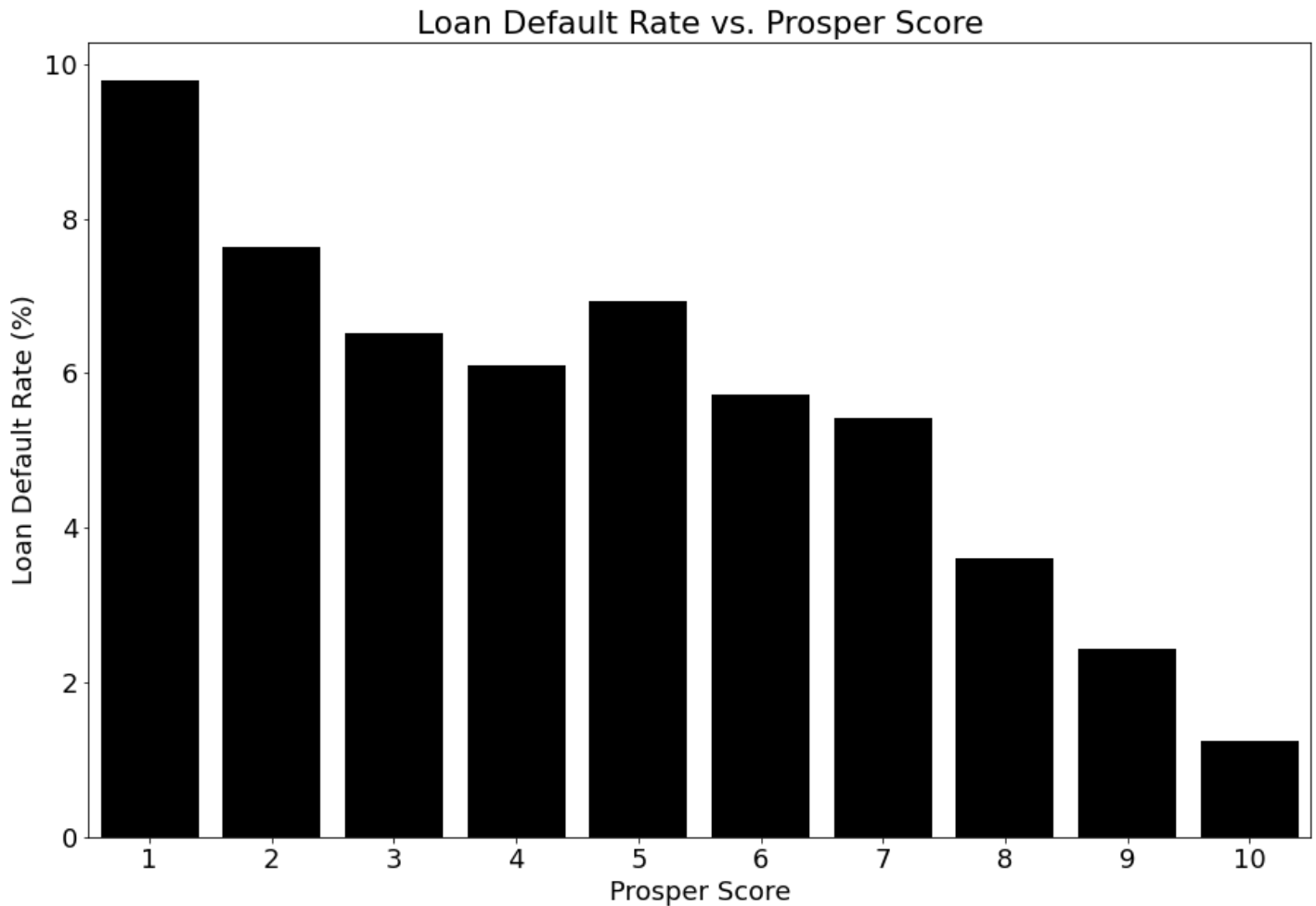
Loan Default vs. Debt Ratio

The loan default rate has a positive relationship with the debt ratio. That is, the default rate is high among clients with high debt-to-income ratios. From the analysis, we found that the default rate almost tripled from 9.4% for clients with low debt ratios (0-13%) to 25.3% for those with high debt ratios (1000% plus).

## Loan Default Rate vs. Debt to Income Ratio
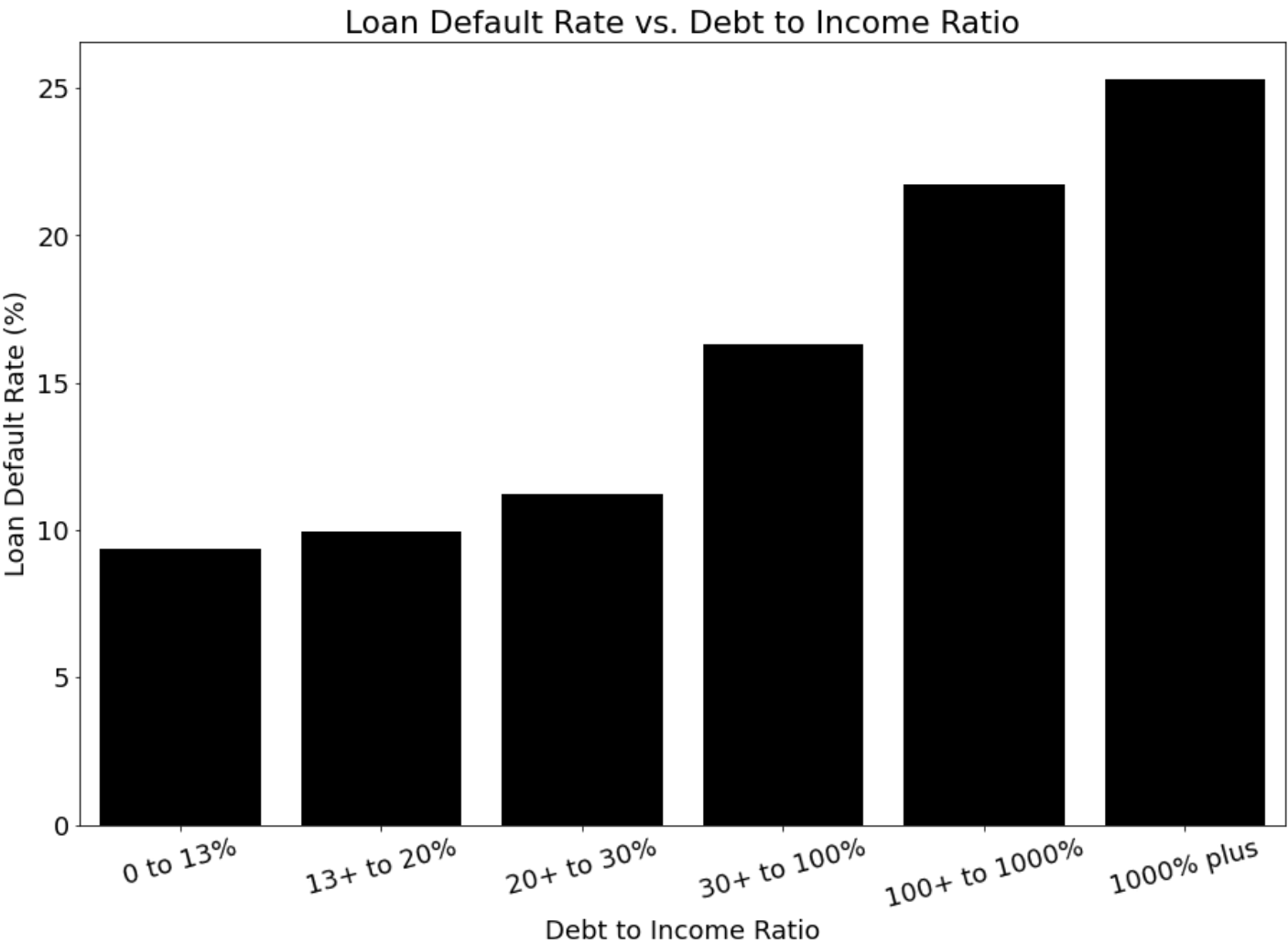
Loan Default vs. Borrower Rate

The analysis reveals that, on average, the borrower rate is higher among clients who defaulted compared to those who completed the loan. The median borrower rate for defaulters is about 22.9% and that of those who completed is 17.5%. This suggests a positive relationship between loan default rates and borrower rates.

Loan Status vs. Borrower Rate
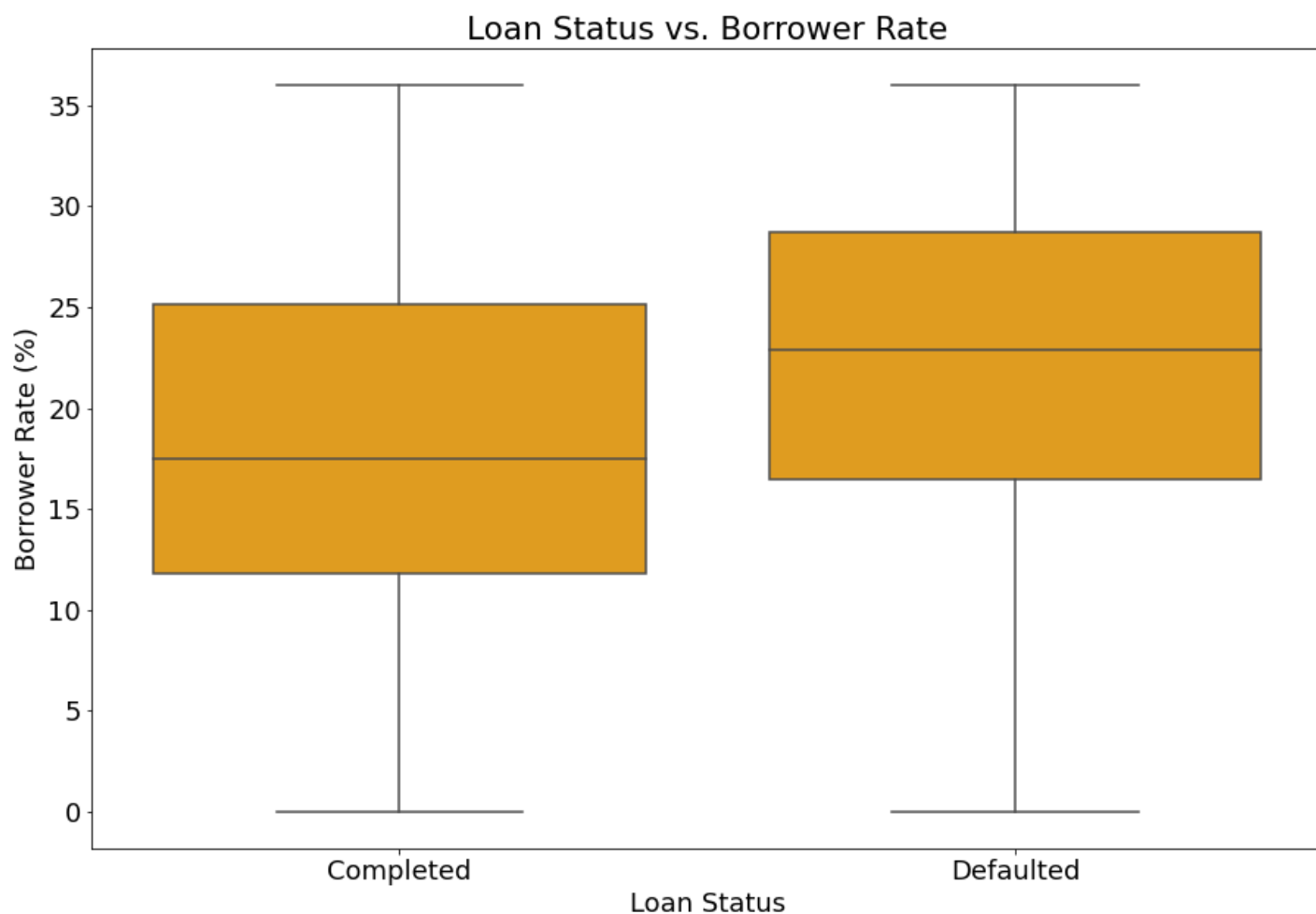
Loan Default vs. Credit Score

On average, Credit Score is lower among clients that defaulted compared to those that completed their loans. Defaulters have a median Credit Score of about 640 compared to 680 for those who completed. This suggests that the loan default rate is negatively related to Credit Score.
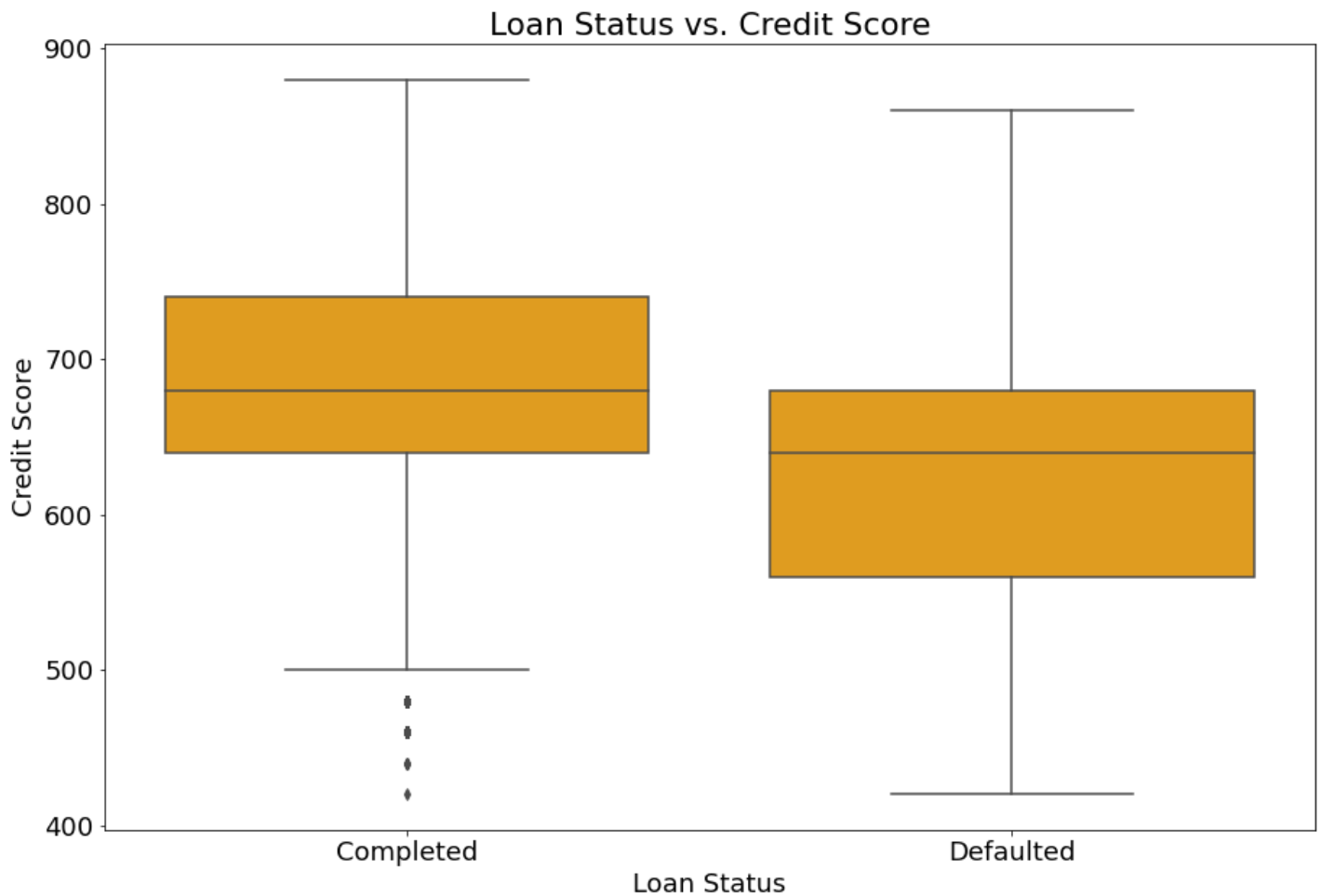
Loan Status and Borrower Rate by Employment Status

Generally, we observe that the borrower rate for clients who defaulted is higher than those who completed the loan. On average, clients in the "Other" employment group have the highest borrower rate. In contrast, those in the "Retired" group have the least borrower rate. Furthermore, notice that the difference in borrower rate between clients that defaulted and those who completed is greatest among clients whose employment status is Not available (8%) and smallest for the Self-employed (0.26%).
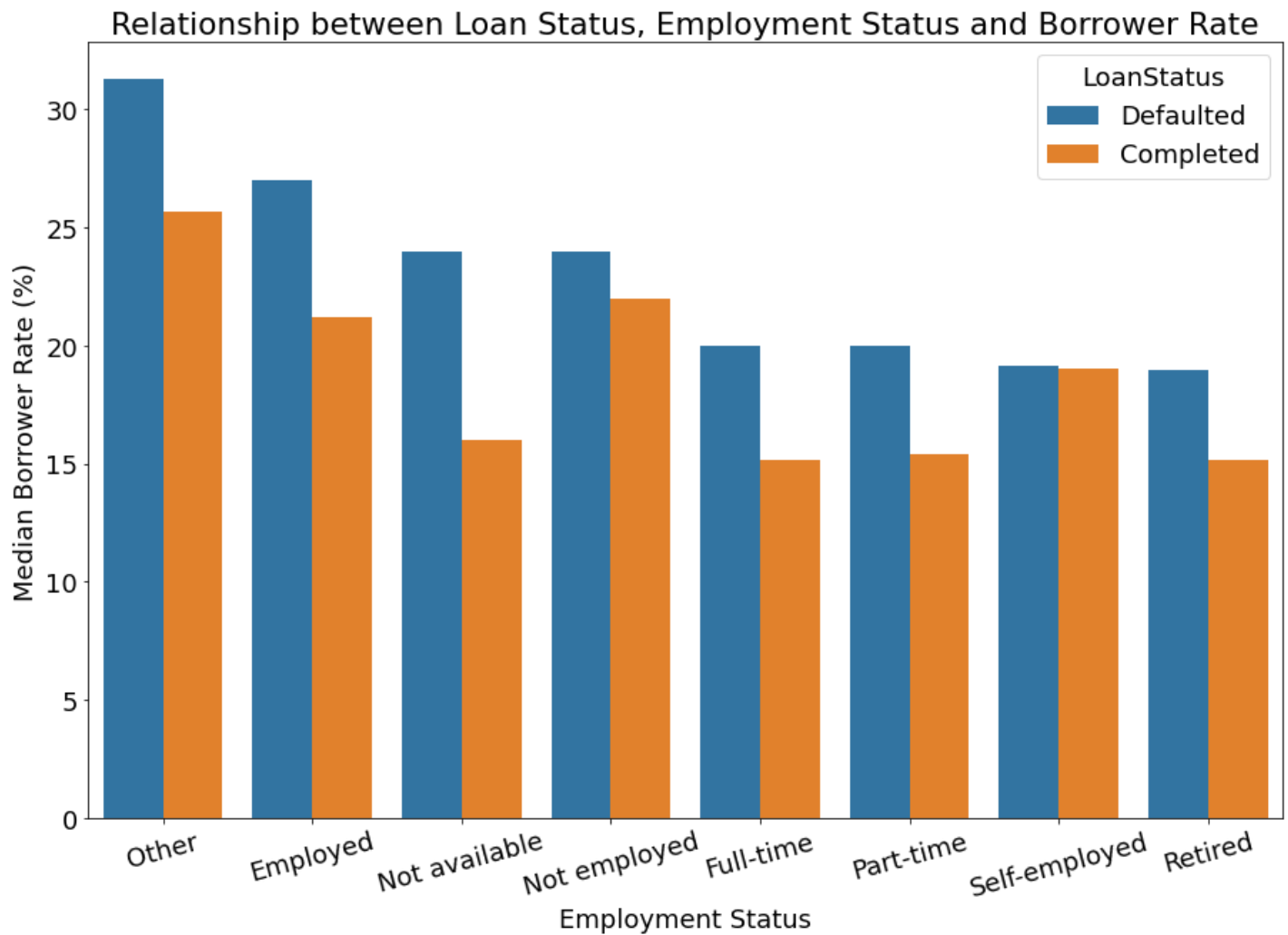
Loan Status and Debt Ratio by Income

On average, for clients whose income is displayed, the debt ratio is higher for clients that defaulted compared to those that completed the loan. Also, the difference in debt ratio between those that defaulted and those that completed is largest among clients who are not employed. For clients whose income level is not displayed, the debt ratio is similar for both loan defaulters and "completers".

Relationship between Loan Status, Income and Debt Ratio
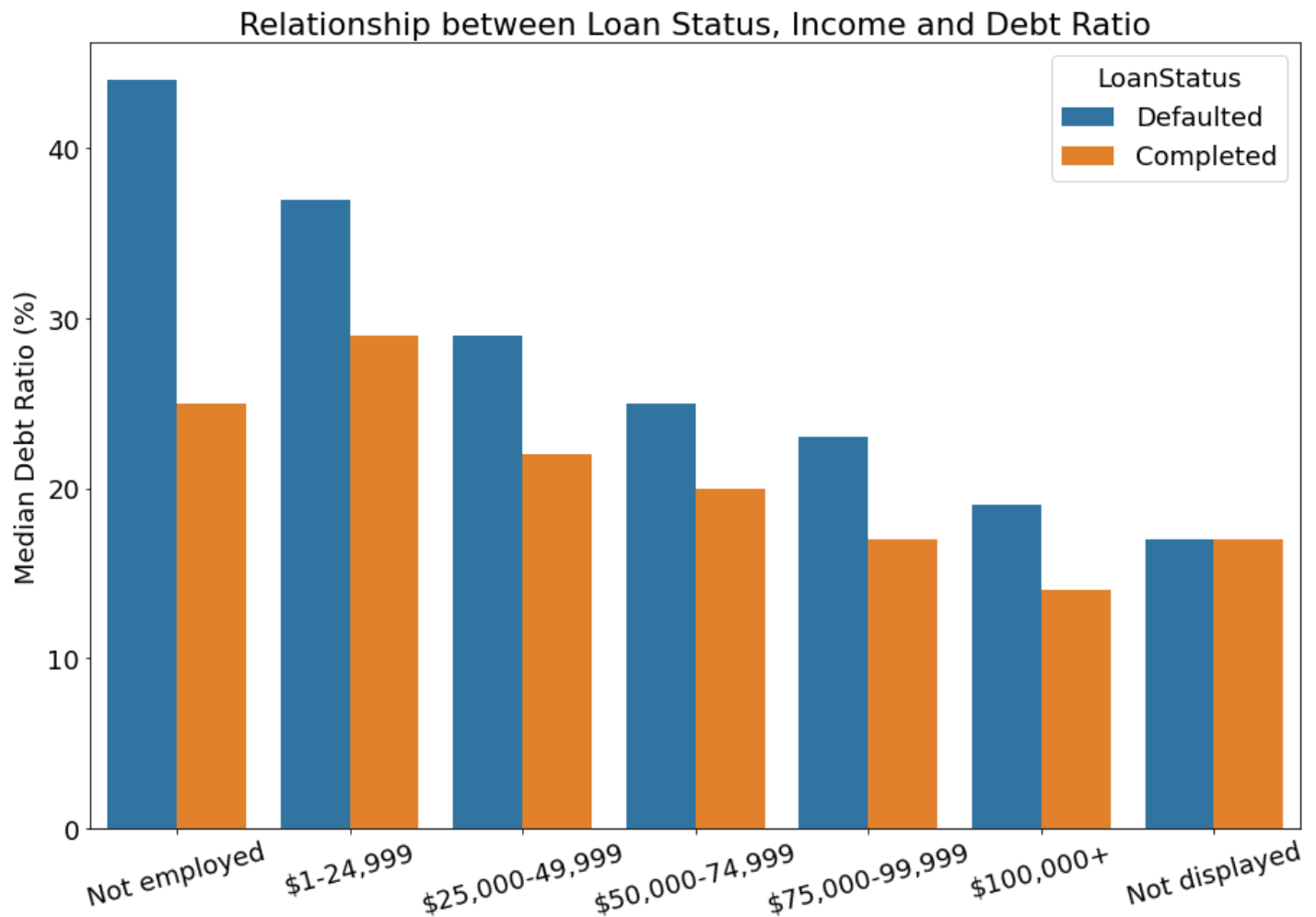
Summary of Key Findings

The key findings are outlined as follows:

- High-income levels correspond to lower loan default rates.
- High Prosper Scores correspond to lower levels of loan default rates.
- Default rate is high among clients with high debt-to-income ratios.
- On average, the borrower rate (interest rate) is higher among clients who defaulted compared to those who completed the loan.
- On average, Credit Score is lower among clients that defaulted compared to those that completed their loans.
- The borrower rate for clients who defaulted is higher than for those who completed the loan.
- On average, for clients whose income is displayed, the debt ratio is higher for clients that defaulted compared to those that completed the loan.