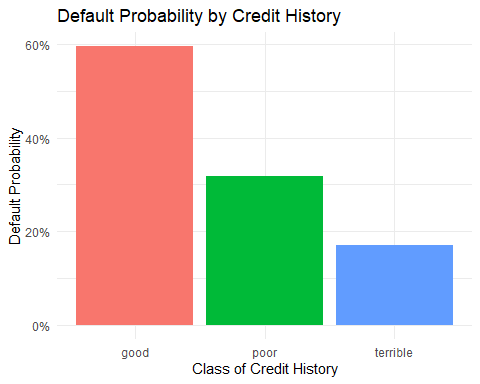
q2.hw2

2024-02-23



## (Intercept) duration amount installment   
## -0.71 0.03 0.00 0.22   
## age historypoor historyterrible purposeedu   
## -0.02 -1.11 -1.88 0.72   
## purposegoods/repair purposenewcar purposeusedcar foreigngerman   
## 0.10 0.85 -0.80 -1.26

The data and analysis from a German bank highlight a puzzling relationship between credit history and loan defaults. The bar plot indicates that worse credit history is associated with higher default rates. However, the logistic regression model suggests the opposite, with ‘poor’ and ‘terrible’ credit histories associated with lower default probabilities, which contradicts common financial intuition.

This contradiction is likely due to the bank’s sampling method, which oversamples defaults. This method creates a bias, making the model less applicable to the general population of borrowers. For the model to be useful in screening new borrowers, the bank needs to use a sample that accurately reflects the true default rate in its overall portfolio. Correcting this sampling bias is essential for the bank to develop a reliable model for classifying borrowers by default risk.