

**Department of Statistics  
The Wharton School  
University of Pennsylvania**

## **Credit Risk Dataset Overview**

### **BUSINESS OVERVIEW**

The lender attracts borrowers that typically have started a small local business, such as a dry cleaners, gasoline station or restaurant and do not have an extensive credit background. Opening this business has stretched the limits of the merchant, and they come to the lender seeking cash in order to keep their business going or to get past a slack period. Their lack of background makes it hard for these entrepreneurs to obtain a line of credit from the local bank, and so they have sought a short-term carry-over loan.

The term of the loans considered in this analysis is 12 months. The principal of the loan ranges from about \$10,000 to more than \$100,000. To obtain payments from the merchant, this lender garnishes the stream of credit card transactions processed by the merchant. Each month, the lender extracts a *fixed* percentage of the value of credit card transactions made by the business. This percentage is set at the time that the loan originates. The lender will extract this percentage from the credit card stream until the loan is repaid (assuming that the business continues to operate). The percentage garnished each month is intended to pay off the loan in 12 months, but that does not always happen because of variation in the flow of transactions.

For example, suppose a borrower operates a restaurant and handles \$30,000 in credit card transactions in a typical month. This borrower obtained a loan of \$25,000. With fees and interest of, say, 20%, the borrower owes the lender the principal \$25,000 plus \$5,000, or a total payment of \$30,000. In order to obtain repayment in 12 months, the lender garnishes the credit card transactions at a rate that should cover this debt in 12 months. To obtain the \$30,000 owed, the lender wants to get  $\$30,000/12 = \$2,500$  per month. On average, then, the lender would plan to extract about 8.33% of the value of the credit card transactions made by this borrower. Of course, the value of transactions varies from month-to-month, so a fixed percentage garnish generates a varying level of repayments monthly. There is random variation in the operations of these borrowers, and hence variation – and risk – in the payments received by the lender. If the borrowing business is doing better than usual (*i.e.*, generating more than \$30,000 in credit card transactions monthly), then the loan will be repaid in less than 12 months. If the business is doing poorly, the lender will have to recover the loan over a longer term.

### **DATA**

The data includes cases and variables that the lender is able to access easily. These variables include the principal and total value (principal plus interest and fees) for

each loan as well as the amount that has been repaid on the loan at 6 months. The data also include potentially useful explanatory variables that describe the borrower, the business operated by the borrower, and demographic information that describes where the business operates. The demographic information gives aggregate characteristics of the postal zip code in which the merchant does business, including population and income. The following table lists all of the variables that are in the data table and gives a concise definition of each.

Variable Name	Description
Case Number	Unique loan identifier.
PRSM	The Performance Ratio at Six Months, defined as twice the ratio of Amount repaid at 6 months to Total amount to be repaid.
Amount repaid at 6 months	The dollar value of payments to the lender after six months.
Nominal loan amount	The amount originally loaned to the borrowing merchant (principal).
Total amount to be repaid	The amount in dollars that the merchant must pay back in order to satisfy terms of the loan
Repayment percentage	Percentage of each month's credit card transactions that will be debited to pay back the loan.
Commission Upfront	The commission paid to the independent sales organization that recruited the borrower.
Validated monthly batch	Average monthly credit card stream in dollars, calculated over all available historical data, as validated by the lender.
Historical monthly credit card receipts	The average monthly credit card stream in dollars, calculated over all available historical data
Months of history	The number of months history that the historical monthly mean was calculated from
Loan type	Either a first [O for original] loan made to this merchant or renewal [R] loan.
Loan size class	A categorical variable indicating the total repayment amount of the loan: S1 = \$0-\$9,999, S2 = \$10,000 - \$24,999 S3 = \$25,000 - \$49,999, S4 = \$50,000+
FICO	Fair-Isaac credit score of the individual guaranteeing the loan
Years In business	The number of years that the merchant has been in business
Num of credit lines	The number of new and continuously reported lines of credit available for this business.
Satisfied Accounts	Total number of credit lines that are now current or paid satisfactorily
Num of paid-off credit lines	Number of credit lines (loans) that the merchant has paid off satisfactorily.
Current delinquent credit lines	Number of current delinquent or derogative lines of credit
Previous delinquent credit lines	Number of previous delinquent or derogative credit accounts
Business entity type	Legal entity, such as "corporation" or "sole

	proprietor”.
Years in file	Number of years the business has had its credit tracked by the credit bureau.
Num of legal items	Number of legal items reported for merchant.
Num of derogatory legal items	Number of federal, state and local tax liens, bankruptcies and judgments on the business.
Num of trade lines	Number of new and continuously reported trade lines for this business.
Two digit SIC code	The type of business, numerically recorded as a two-digit integer.
Two digit SIC description	Text description of the type of business
Population in zip code	The population in the zip code where the merchant does business.
Male population in zip code	Number of males in the zip code where the merchant does business (thousands)
Female population in zip code	The number of females in the zip code where the merchant does business
Average house value in zip code	The average house value in the zip code where the merchant does business
Income per household in zip code	The income per household in the zip code where the merchant does business (recorded in dollars).
Median age in zip code	The median age of people living in the zip code where the merchant does business
Time zone	The time zone of the zip code where the merchant does business (hours relative to Greenwich).
Bus establishments in zip code	The number of business establishments in the zip code where the merchant does business
Employment in zip code	The number of persons employed in the zip code where the merchant does business
Annual payroll in zip code	The annual gross payroll in the zip code where the merchant does business
PO residential count in zip code	The number of residential delivery mail stops in the zip code where the merchant does business
PO business count in zip code	The number of business delivery mail stops in the zip code where the merchant does business
Delivery total	The total number of mail stops in the zip code where the merchant does business
State	Two-letter postal code for the state in which the merchant operates.
ISO Name	Abbreviated name of the independent sales organization that originated the loan.