VICTORIA FADEKEMI OJO

ISOM 837

PROFESSOR KIM SEOKJIN

**STORYTELLING PREP REPORT**

I will be making use of loan data set which is from LendingClub.com and it spans the years 2007-2010 and involves the classification and prediction of whether borrowers paid back their loans in full. It contains about 14 variables and about 9,579 rows. Listed below are the summary of the columns:

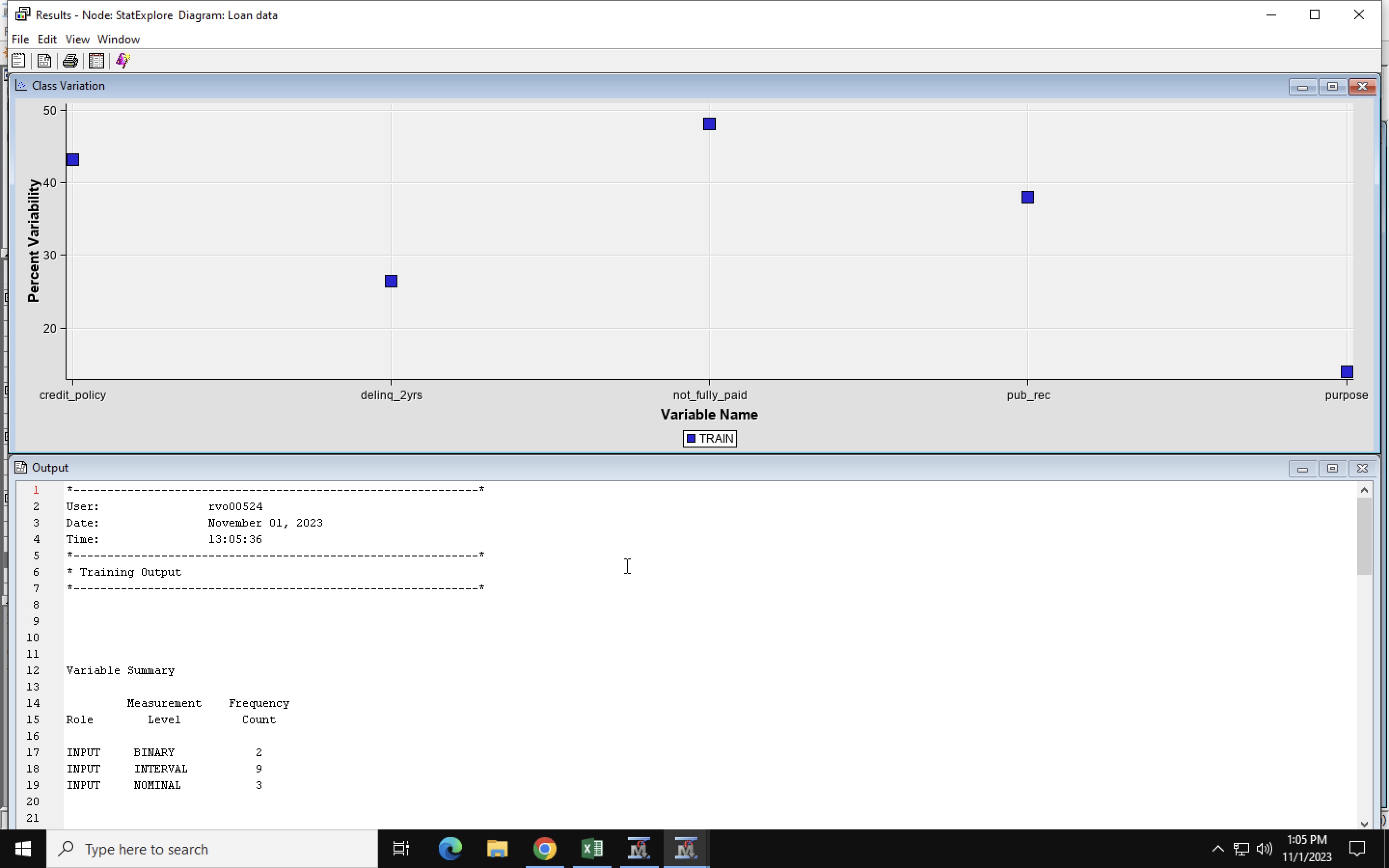
* **Credit. Policy**: A binary variable that indicates if a consumer satisfies the requirements for credit underwriting set out by LendingClub.com (1 if the requirements are met, 0 otherwise).
* **Purpose**: This is broken down into categories like "credit card," "debt consolidation," "educational," "major purchase," "small business," and "all other."
* **Int.rate**: A proportional representation of the loan's interest rate. The interest rates for riskier borrowers are higher.
* **Installment**: Should the loan be funded, the borrower's monthly payment obligations.
* **Log.annual.inc**: The natural log of the borrower's self-reported yearly income. DTI stands for debt-to-income ratio, which is computed by dividing total debt by yearly income.
* **FICO:** This stands for the borrower's FICO credit score.
* **The days.with.cr.line** : How long has the borrower had access to a credit line?
* **Revolving balance (revol.bal)**: The amount owed by the borrower at the conclusion of each credit card payment cycle.
* **Revol.util**: Revolving line utilization rate, which shows how much of a credit line is really used in comparison to the total amount of credit available.
* **Inq.final.6 months**: This shows the amount of creditor inquiries the borrower has had in the previous six months.
* **Delinq.2yrs**: The total number of instances during the previous two years in which the borrower has missed a payment by more than thirty days.
* **Pub.rec:** The quantity of adverse public documents pertaining to the borrower, such as judgments, tax liens, or bankruptcy filings.
* **Not Fully Paid**: This simply means that the loan was not repaid in whole, meaning that the borrower didn't follow the terms of repayment in full.

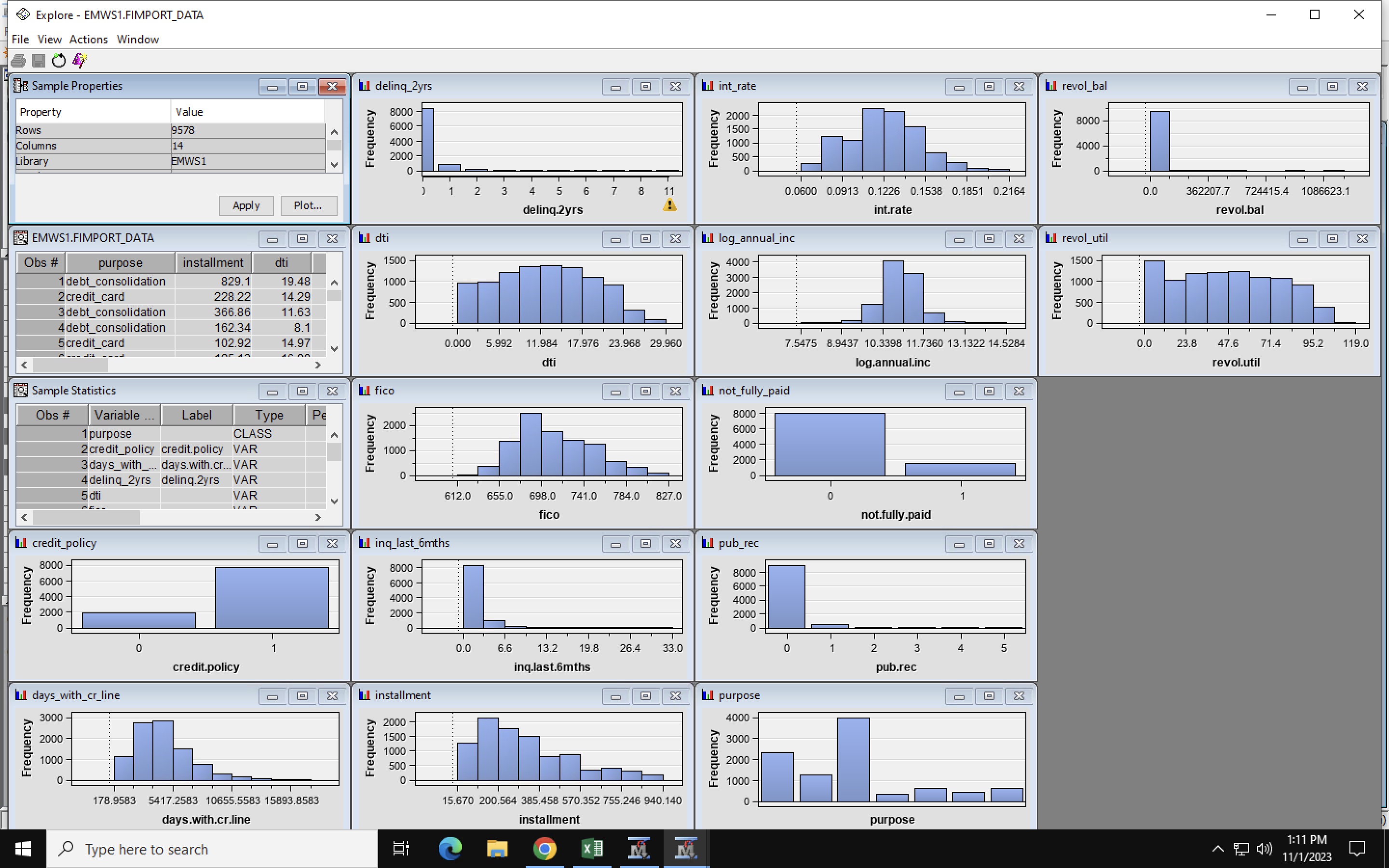
**POSSIBLE RESERACH QUESTIONS**

**“Not fully paid” will be my target variable.**

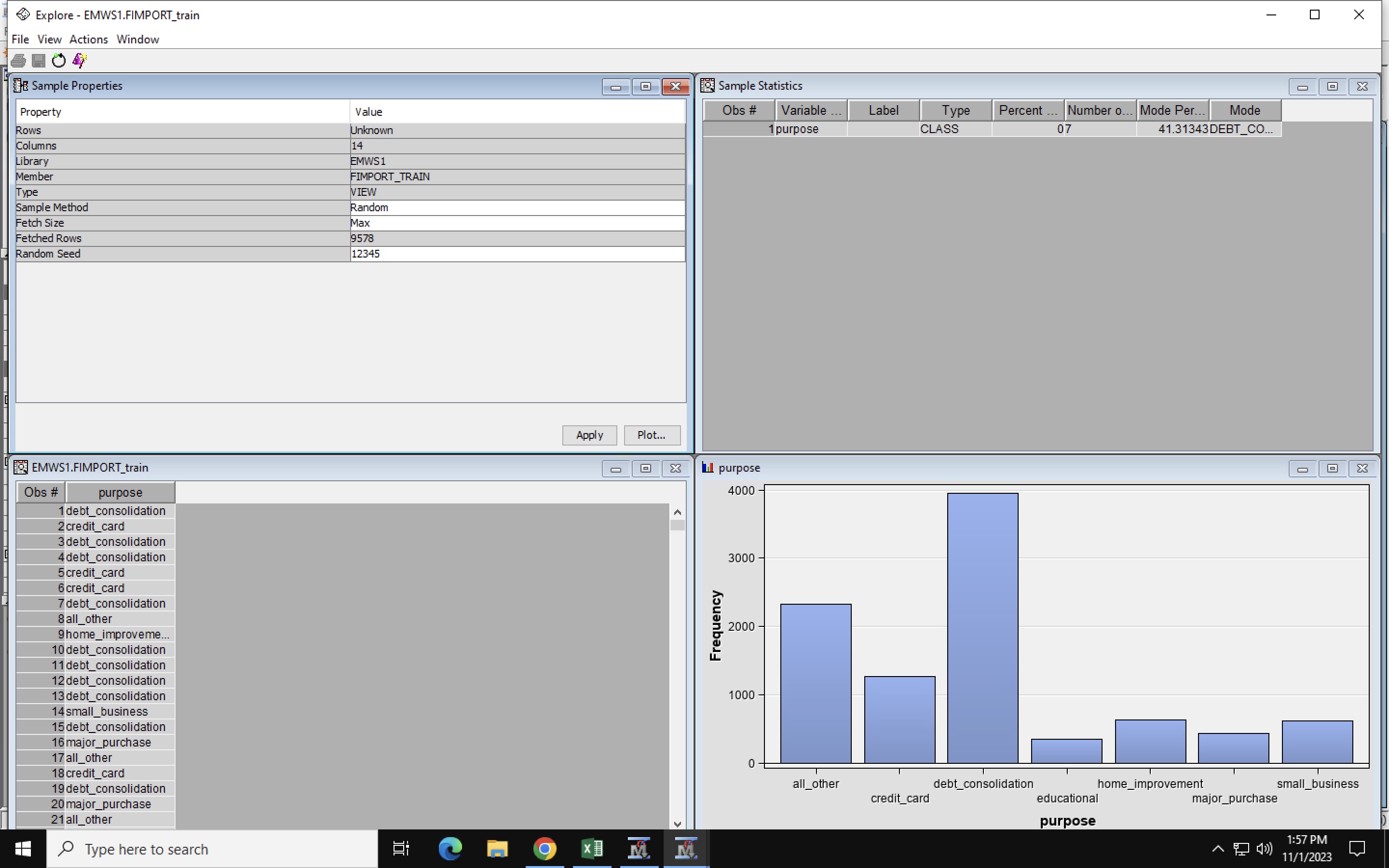
* How does the loan's purpose affect the chance that it won't be repaid in full?
* Are those who borrow money more likely to default on their debts if their natural log of annual income (log.annual.inc) is lower?
* What is the relationship between the frequency of non-payment and the number of delinquencies in the last two years (delinq.2yrs)?

**Summary Statistics;**



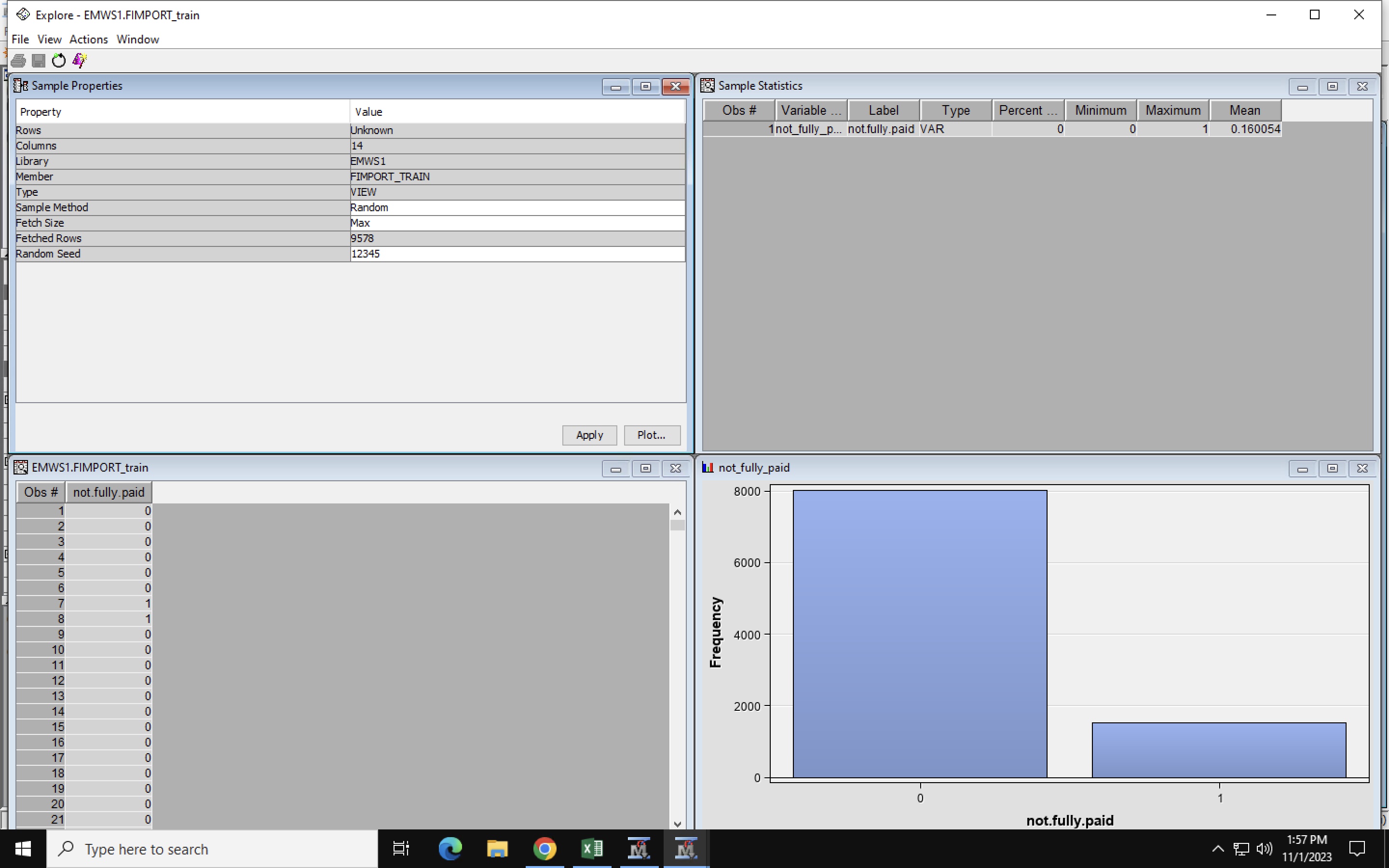


**Exploring the loan purpose data set.**



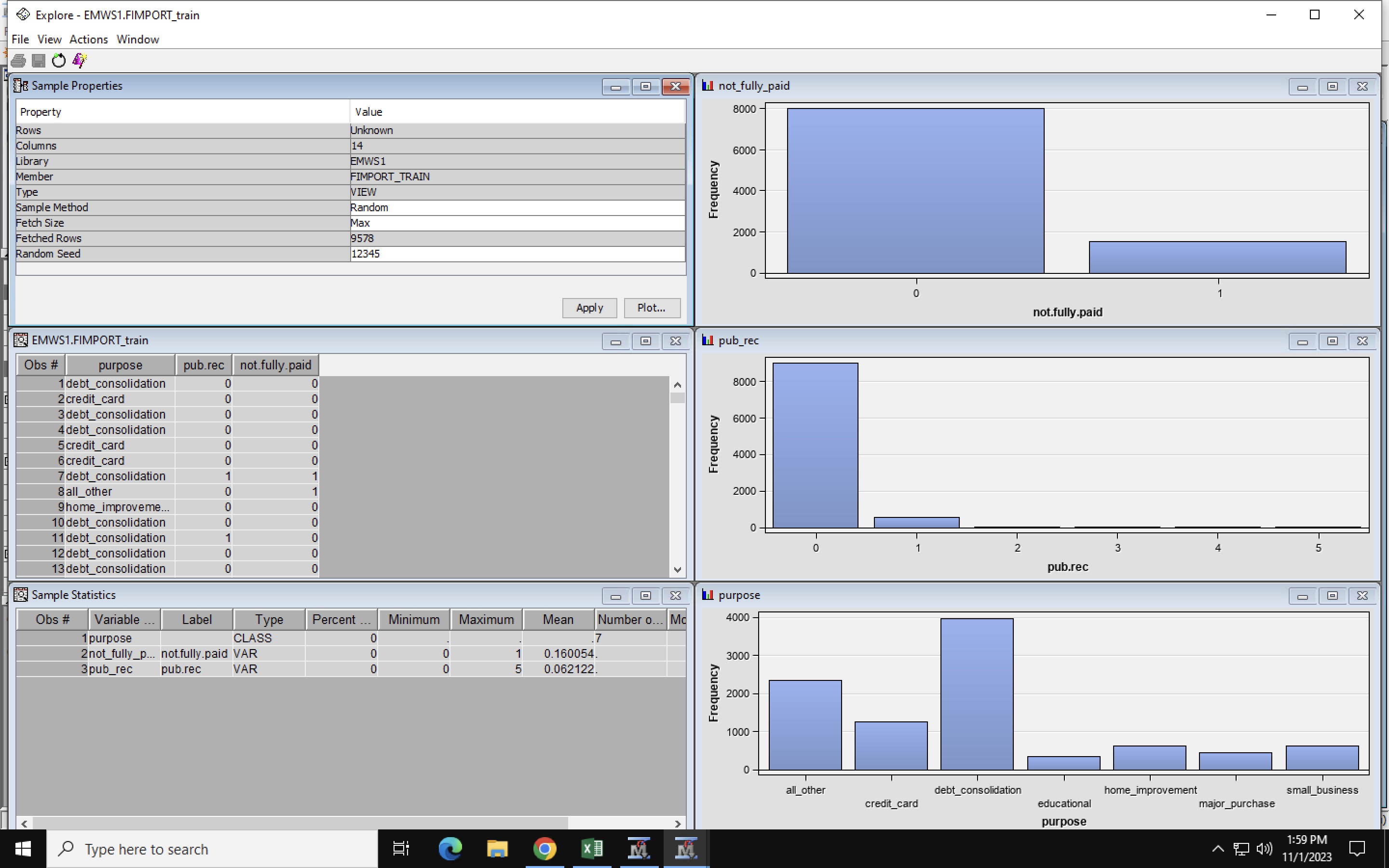
We can see that the highest purpose for loan is for debt consolidation.

**Exploring “not fully paid” data set.**



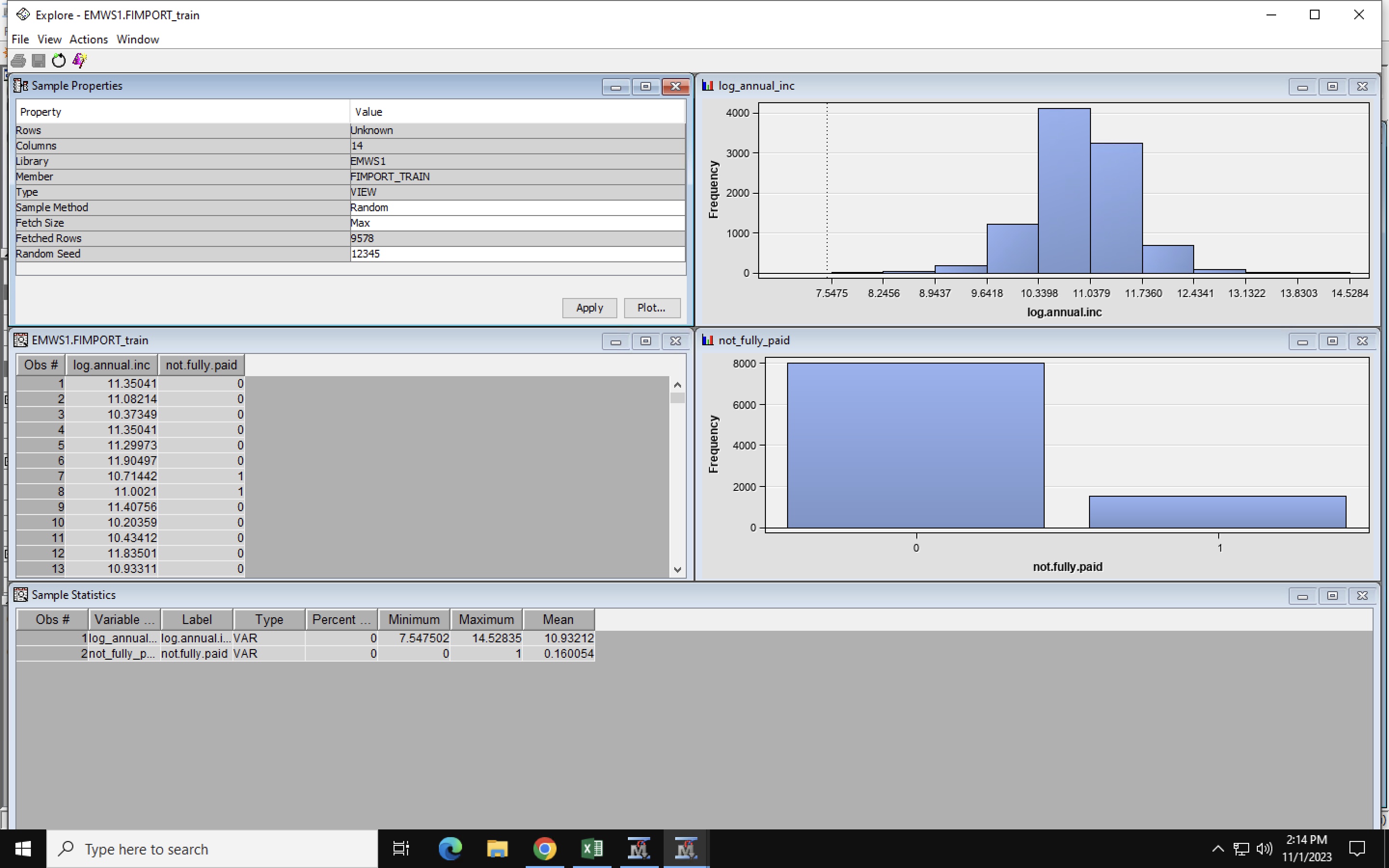
**About 8,045 individuals have not fully paid their loans.**

**Only 1,533 have fully repaid their loan.**



Debt Consolidation has the highest number of individuals who have not fully repaid their loan.

**Repayment & Log annual inc.**



**This shows that individuals who earn between 10.33984 and 11.03793 have not fully repaid their loan.**

**It comes up to a total of 4123.**