

Project2 - Part 1

1. Team members – Max 2

Alice Jiayi Wang

Zekun Wang

2. Describe your dataset? Where are you getting it from? What is the data about? Size? How many attributes? (Remember it should be comparative to voter data for project 1)

We found it from Kaggle. It has the complete loan data for all loans issued through the 2007-2015, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information. The file containing loan data through the "present" contains complete loan data for all loans issued through the previous completed calendar quarter. Additional features include credit scores, number of finance inquiries, address including zip codes, and state, and collections among others.

The file is a matrix of about 890 thousand observations and 75 variables.

3. What are you planning to do in your Project 2? High level goals? What is your application going to be?

We are planning to build a dashboard based on this dataset.

High level goals are connecting the database functionality to the application interface, as well as enabling user interaction with the underlying data.

4. What are your next steps? Deliverables in next step? In a week time what do you plan to do?

Understand the meaning for each variable and find the dependencies between them.

The deliverables in next step would be a UML database.