**Data size**: 141740

**Attributes size**: 81.

Attributes’ definitions and my thinking about every attribute (Keep or Drop, the value range about every attributes, and the mistakes about some attributes) are in the **LoanDataExport-DataDictionary.xlsx.**

**General methodology:**

1. Since there are many missing values in the dataset, we do not drop all of them. We only drop the missing values when we want to analyze a specific attribute. For instance, if we want to analyze the employment status, we drop all the missing values in employment status, but do not drop other attribute’s missing values. If we want to analyze the relation between borrower states and employment status, we only drop all the missing values in these two attributes.
2. Classify the attributes into some categories. And analyze their priorities.

**Drop attributes: (24)**

Listing key

Listing number

Listing creation date (use loan origination date)

Currently in group

Group key

Open revolving monthly payment

Public record last 10 years

Public records last 12 months

Available bank card credit

Total trades

Trades never delinquent

Trades opened last 6 months

Income verifiable

Loan key

Loan first defaulted cycle number

Loan number

Loan origination quarter

Member key

Monthly loan payment

Percent funded

Recommendations

Investment from friends count

Investment from friends amount

Investors

**1st priority attributes: (Bold + Red) (22)**

Credit grade/ Prosper rating

Term

Loan status

Closed date

Borrower APR

Borrower Rate

Category

Occupation

Employment status

Employment status duration

Is borrower homeowner

Credit score range lower

Credit score range upper

First record credit line

Inquires last 6 months

Revolving credit balance

Bankcard utilization

Debt to income ratio

Income range

Stated monthly income

Loan original amount

Loan origination date

**2nd priority attributes (Bold) (13)**

Lender Yield

Borrower state

Date credit pulled

Current credit lines

Open credit lines

Total credit lines past 7 years

Open revolving accounts

Total inquires

Current delinquencies

Amount delinquent

Delinquencies last 7 years

Loan current days delinquent

Loan months since origination

**Attributes only occurred after July 2009** **(Underlined) (there are 112656 data after July 2009) (6)**

Estimated effective yield

Estimated loss

Estimated return

Prosper rating (numeric)

Prosper rating (alpha)

Prosper score

**Attributes only occurred when the borrower has prior loans (Underlined) (23884 data have prior loans) (8)**

Total prosper loans

Total prosper payments billed

On time prosper payments

Prosper payments less than one month late

Prosper payments one month plus late

Prosper principal borrowed

Prosper principal outstanding

Score change at time of listing

**Attributes to analyze charged off/defaulted loans: (Underlined) (11636 charged off, 6212 defaulted) (8)**

LP\_Customer payments

LP\_Customer principal payments

LP\_Interestand Fees

LP\_Service Fees

LP\_Collection Fees

LP\_Gross principal Loss

LP\_Net principal loss

LP\_Non principal recovery payments