Data Analysis

Mission

1. How to you ensure that customers can/will pay their loans.
2. Can CreditOne approve loans with high certainty

Data used for Analysis

Historical data of customers with following information are provided

Dataset statistics

|  |  |
| --- | --- |
| **Number of variables** | 25 |
| **Number of observations** | 30000 |
| **Missing cells** | 0 |
| **Missing cells (%)** | 0.0% |
| **Duplicate rows** | 0 |
| **Duplicate rows (%)** | 0.0% |

Limit\_Bal is a continuous variable with following values

|  |  |
| --- | --- |
| **Minimum** | 10000 |
| **Maximum** | 1000000 |

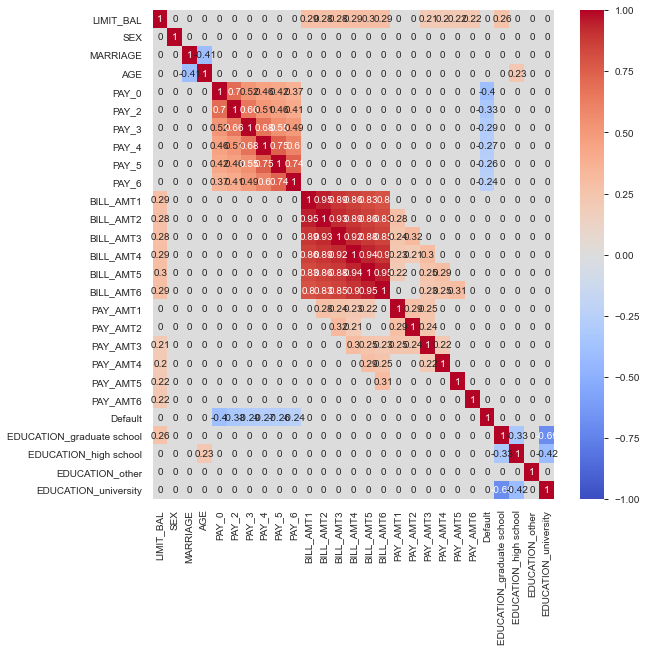
[LIMIT\_BAL](file:///C:\Users\jvarughe.JCV20\output.html#pp_var_-333569038476698593)  
Real number (ℝ≥0)

|  |  |
| --- | --- |
| **Distinct count** | 81 |
| **Unique (%)** | 0.3% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 167442.00500584015 |
| **Minimum** | 10000 |
| **Maximum** | 1000000 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Memory size** | 234.2 KiB |

Discretization is the process of transforming a continuous-valued variable into a discrete one by creating a set of contiguous intervals (or equivalently a set of cutpoints) that spans the range of the variable’s values.

Data mining objectives

1. Decrease customer default rates
2. Increase revenue by limiting loans which can be paid back.
3. Understand the correlation between given features and loan default



Using the discretization method to improve predictability

