**Explanation for Bene1cont data set**

Number of observations: 3123

Number of good customers: 2082 (66.67%)

Number of bad customers: 1041 (33.33%)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Number** | **Name** | **Type** | **Explanation** | **Number of different values ​​according to data dictionary** | **Number of different values ​​in data set** | **Note** |
| 1 | NUMLENI | continuous | loan number |  |  |  |
| 2 | BLENIA1 | continuous | amount of the loan |  |  |  |
| 3 | BFACTA1 | continuous | invoiced amount |  |  |  |
| 4 | PLASTA1 | continuous | burden percentage |  |  |  |
| 5 | PDUURA2 | continuous | duration in months |  |  |  |
| 6 | CDOELA2 | nominal | target loan officially |  | 34 |  |
| 7 | CGEBRA1 | nominal | code usage  1 = private, 2 = professional use | 2 | 2 |  |
| 8 | BMENSA1 | continuous | humanity |  |  |  |
| 9 | BSPARA11 | continuous | save amount balance |  |  |  |
| 10 | BUITGA21 | continuous | expenses other credit expenses |  |  |  |
| 11 | BINKOA11 | continuous | professional income |  |  |  |
| 12 | DBEGIA21 | continuous | Number of years in service at the time of application |  |  |  |
| 13 | DVERBA11 | continuous | Number of years the client stays in Belgium at the time of application |  |  |  |
| 14 | DGEBOA11 | continuous | age at time of application |  |  |  |
| 15 | CPRIVA11 | nominal | private person code  1 = employee  2 = self-employed  3 = non-private person  4 = without occupation | 4 | 2 |  |
| 16 | CBURGA11 | nominal | Civil status code | 7 | 5 |  |
| 17 | DVERBA21 | continuous | Number of years since last move |  |  |  |
| 18 | CTROSA11 | nominal | code loyal saver | 3 | 2 |  |
| 19 | CEIGEA11 | nominal | the person concerned owns a property | 7 | 4 |  |
| 20 | DCLIEA11 | continuous | Number of years of client at the time of request |  |  |  |
| 21 | DLLENA11 | continuous | number of years since last loan at the time of application |  |  |  |
| 22 | ACONTCV1 | continuous | Number of current accounts |  |  |  |
| 23 | ACONTTM1 | continuous | Number of term accounts |  |  |  |
| 24 | ACONTHY1 | continuous | Number of mortgage loans |  |  |  |
| 25 | APERSA11 | continuous | Number of dependents |  |  |  |
| 26 | CECOTA11 | nominal | code of the economic sector of employment of the person concerned | 8 | 8 |  |
| 27 | CVOLTA11 | nominal | code that indicates whether the person is employed full-time or not | 7 | 7 |  |
| 28 | CAANSA11 | nominal | code that indicates the title | 6 | 3 |  |
| 29 | CLASS |  |  |  |  |  |

**Some tips when using this data set**

The attribute no. 7 ( **CDOELA2**) contains 34 different values.

Joseph Vingerhoets proposed the following recoding:

|  |  |  |
| --- | --- | --- |
| 25 - 28 and 70 | cars new | 1 |
| 23 - 24 | cars up to 2 years | 2 |
| 22 and 71 | used cars | 3 |
| 91 - 92 | immo | 4 |
| 89, 99 | treasury | 5 |
| 30-34 and 36-39 and 41-45 | objects | 6 |
| 51 - 59 | holidays | 7 |
| 11 - 13 | motorbikes | 8 |
| 35 and 60-69 and 93 | everything in house | 9 |
| 40 and 50 | courses and travel | 10 |
| 10, 20, 21, 96, 98 | household expenses | 11 |
|  |  |  |

It is especially useful to keep the first 6 classes separate. You can still combine the other according to your wishes.

Reasons:

cars new = small risk

used cars = real risk of fraud

cars up to 2 years = something in between

immo = small number of loans, but there are more large amounts

Treasury = the large category of loans that are not included for what they are used for = reasonable risk

I have therefore recoded the CDOELA2 attribute!

The **CPANDA11**attribute (attribute no. 30) has a strange coding. Take this into account! Maybe possibly leave out.

The **CBEROA21**attribute (appeal code, attribute no. 13) is partly redundant with the CECOTA11 attribute (economic sector employment code of the person concerned).

We still have to check the coding of the **CNATIA21**attribute (attribute no. 18).

The following attributes are omitted from the original input.txt file:

No. 6. CPPLFA1

No. 13 CBEROA21

No. 18 CNATIA21

No. 23 CKINFA11

No. 30 CPANDA11