

training__dataset

Autogenerated data summary from dataMaid

2018-10-24 00:12:25

Part 1

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	27000
Number of variables	24

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×				
Identify levels with < 6 obs.	×	×	×				
Identify case issues	×	×	×				
Identify misclassified numeric or integer variables	×	×	×				
Identify outliers				×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

Part 2

Summary table

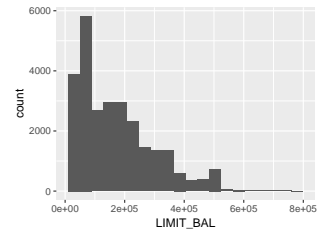
	Variable class	# unique values	Missing observations	Any problems?
LIMIT_BAL	numeric	80	0.00 %	×
SEX	integer	2	0.00 %	
EDUCATION	integer	7	0.00 %	×
MARRIAGE	integer	4	0.00 %	
AGE	integer	55	0.00 %	×
PAY_0	integer	11	0.00 %	×
PAY_2	integer	10	0.00 %	×
PAY_3	integer	11	0.00 %	×
PAY_4	integer	11	0.00 %	×
PAY_5	integer	10	0.00 %	×
PAY_6	integer	10	0.00 %	×
BILL_AMT1	numeric	20755	0.00 %	×
BILL_AMT2	numeric	20403	0.00 %	×
BILL_AMT3	numeric	20099	0.00 %	×
BILL_AMT4	numeric	19687	0.00 %	×
BILL_AMT5	numeric	19183	0.00 %	×
BILL_AMT6	numeric	18835	0.00 %	×
PAY_AMT1	numeric	7444	0.00 %	×
PAY_AMT2	numeric	7397	0.00 %	×
PAY_AMT3	numeric	7031	0.00 %	×
PAY_AMT4	numeric	6507	0.00 %	×
PAY_AMT5	numeric	6469	0.00 %	×
PAY_AMT6	numeric	6517	0.00 %	×
default.payment.next.month	integer	2	0.00 %	

Part 3

Variable list

LIMIT_BAL

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	80
Median	140000
1st and 3rd quartiles	50000; 240000
Min. and max.	10000; 8e+05

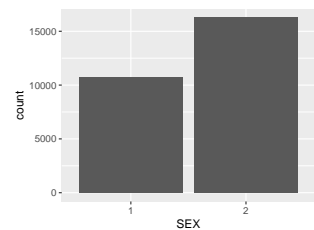


- Note that the following possible outlier values were detected: "740000", "750000", "760000", "780000", "8e+05".

SEX

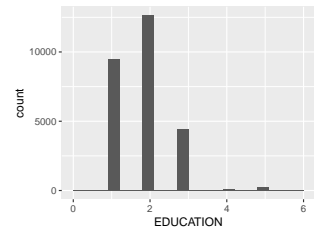
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"2"



EDUCATION

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	7
Median	2
1st and 3rd quartiles	1; 2
Min. and max.	0; 6

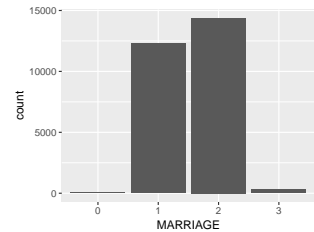


- Note that the following possible outlier values were detected: "3", "4", "5", "6".

MARRIAGE

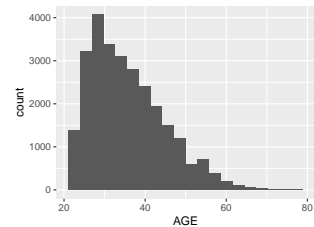
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"2"



AGE

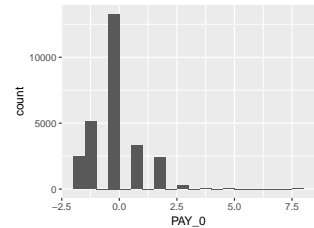
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	55
Median	34
1st and 3rd quartiles	28; 41
Min. and max.	21; 79



- Note that the following possible outlier values were detected: "75", "79".

PAY_0

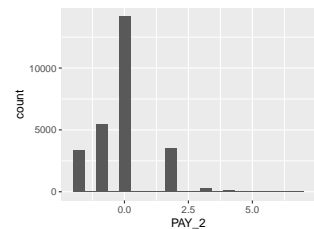
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	11
Median	0
1st and 3rd quartiles	-1; 0
Min. and max.	-2; 8



- Note that the following possible outlier values were detected: "2", "3", "4", "5", "6", "7", "8".

PAY_2

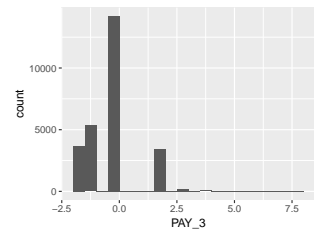
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	10
Median	0
1st and 3rd quartiles	-1; 0
Min. and max.	-2; 7



- Note that the following possible outlier values were detected: "1", "2", "3", "4", "5", "6", "7".

PAY_3

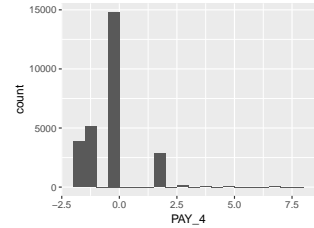
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	11
Median	0
1st and 3rd quartiles	-1; 0
Min. and max.	-2; 8



- Note that the following possible outlier values were detected: "1", "2", "3", "4", "5", "6", "7", "8".

PAY_4

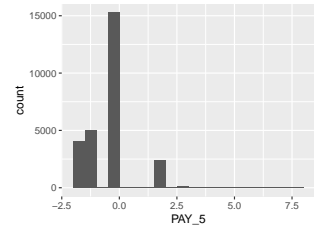
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	11
Median	0
1st and 3rd quartiles	-1; 0
Min. and max.	-2; 8



- Note that the following possible outlier values were detected: "1", "2", "3", "4", "5", "6", "7", "8".

PAY_5

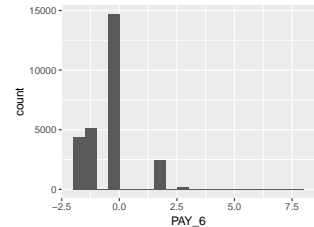
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	10
Median	0
1st and 3rd quartiles	-1; 0
Min. and max.	-2; 8



- Note that the following possible outlier values were detected: "2", "3", "4", "5", "6", "7", "8".

PAY_6

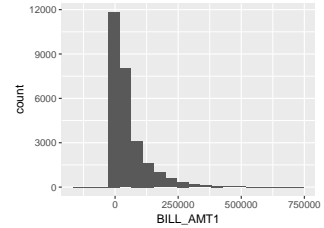
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	10
Median	0
1st and 3rd quartiles	-1; 0
Min. and max.	-2; 8



- Note that the following possible outlier values were detected: "2", "3", "4", "5", "6", "7", "8".

BILL_AMT1

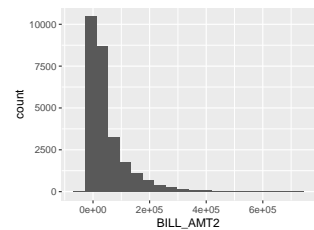
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	20755
Median	22408
1st and 3rd quartiles	3580; 67371.75
Min. and max.	-165580; 746814



- Note that the following possible outlier values were detected: "-165580", "-154973", "-15308", "-14386", "-11545", "-9802", "-9095", "510367", "518950", "519901" (23 additional values omitted).

BILL_AMT2

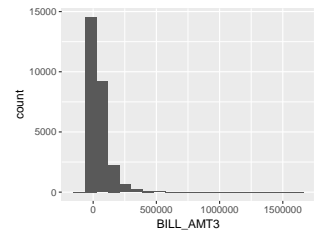
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	20403
Median	21216
1st and 3rd quartiles	2989.75; 64112
Min. and max.	-69777; 743970



- Note that the following possible outlier values were detected: "-69777", "-67526", "-33350", "-30000", "-26214", "-24704", "-22960", "-18618", "-18088", "-17810" (38 additional values omitted).

BILL_AMT3

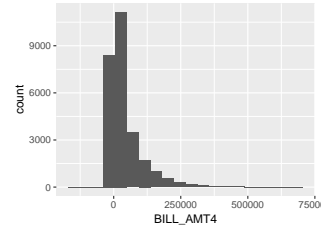
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	20099
Median	20111.5
1st and 3rd quartiles	2683.5; 60385
Min. and max.	-157264; 1664089



- Note that the following possible outlier values were detected: "-157264", "-61506", "-46127", "-34041", "-25443", "-24702", "-20320", "-15910", "-15641", "-15000" (52 additional values omitted).

BILL_AMT4

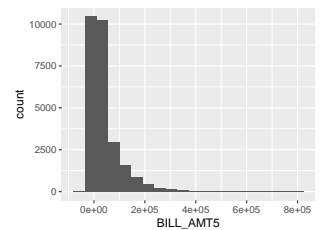
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	19687
Median	19078
1st and 3rd quartiles	2302.5; 54647.75
Min. and max.	-170000; 706864



- Note that the following possible outlier values were detected: "-170000", "-81334", "-65167", "-50616", "-46627", "-34503", "-24303", "-22108", "-20320", "-17250" (92 additional values omitted).

BILL_AMT5

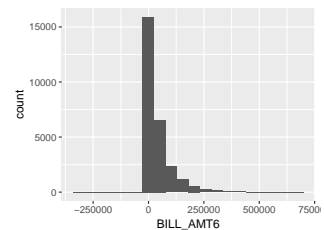
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	19183
Median	18124
1st and 3rd quartiles	1762.75; 50321.25
Min. and max.	-81334; 823540



- Note that the following possible outlier values were detected: "-81334", "-61372", "-53007", "-46627", "-37594", "-36156", "-30481", "-28335", "-23003", "-20753" (146 additional values omitted).

BILL_AMT6

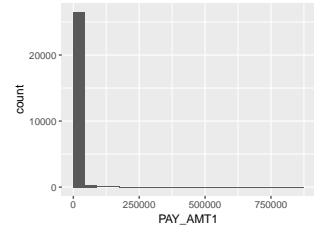
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	18835
Median	17086.5
1st and 3rd quartiles	1244.75; 49198.25
Min. and max.	-339603; 699944



- Note that the following possible outlier values were detected: "-339603", "-209051", "-150953", "-94625", "-73895", "-57060", "-51443", "-51183", "-46627", "-45734" (168 additional values omitted).

PAY_AMT1

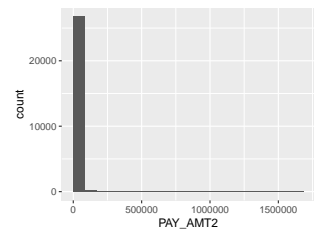
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	7444
Median	2109.5
1st and 3rd quartiles	1000; 5011
Min. and max.	0; 873552



- Note that the following possible outlier values were detected: "0", "1", "2", "3", "4", "5", "6", "7", "8", "9" (752 additional values omitted).

PAY_AMT2

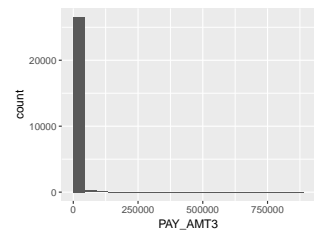
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	7397
Median	2010
1st and 3rd quartiles	820; 5000
Min. and max.	0; 1684259



- Note that the following possible outlier values were detected: "0", "1", "2", "3", "4", "5", "6", "7", "8", "9" (607 additional values omitted).

PAY_AMT3

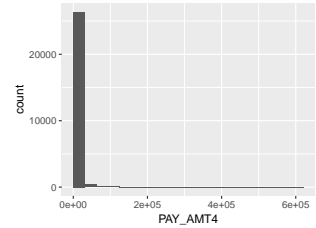
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	7031
Median	1800
1st and 3rd quartiles	390; 4523.25
Min. and max.	0; 889043



- Note that the following possible outlier values were detected: "25448", "25500", "25528", "25630", "25681", "25740", "25753", "25821", "25846", "25854" (738 additional values omitted).

PAY_AMT4

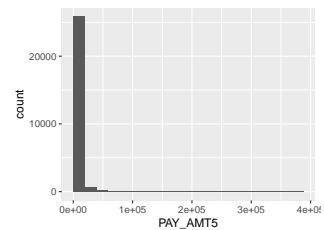
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	6507
Median	1500
1st and 3rd quartiles	300; 4018.25
Min. and max.	0; 621000



- Note that the following possible outlier values were detected: "25896", "25915", "26000", "26004", "26068", "26100", "26130", "26207", "26212", "26249" (679 additional values omitted).

PAY_AMT5

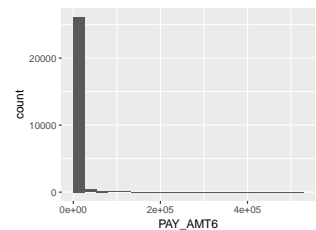
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	6469
Median	1500
1st and 3rd quartiles	248; 4055
Min. and max.	0; 388071



- Note that the following possible outlier values were detected: "25832", "25869", "25910", "25921", "25948", "25949", "25965", "25996", "26000", "26045" (675 additional values omitted).

PAY_AMT6

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	6517
Median	1500
1st and 3rd quartiles	117.75; 4000
Min. and max.	0; 528666

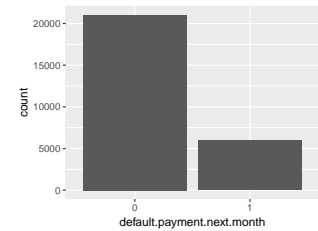


- Note that the following possible outlier values were detected: "25304", "25311", "25349", "25417", "25462", "25497", "25500", "25518", "25586", "25594" (786 additional values omitted).

default.payment.next.month

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"



Report generation information:

- Created by No Instance(s) Available. (username: guzzoan1).
- Report creation time: mer ott 24 2018 00:12:28
- Report Was run from directory: C:/Users/guzzoan1/PythonNotebooks/advml/Assignement1
- dataMaid v1.1.2 [Pkg: 2018-05-03 from CRAN (R 3.5.1)]
- R version 3.5.1 (2018-07-02).
- Platform: x86_64-w64-mingw32/x64 (64-bit)(Windows >= 8 x64 (build 9200)).
- Function call: `makeDataReport(data = training_dataset)`