Loan_approval

Autogenerated data summary from dataMaid

2021-10-24 15:53:11

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	614
Number of variables	12

Checks performed

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

	character	factor	labelled	haven 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.

Summary table

	Variable class	# unique values	Missing observations	Any problems?
Gender	factor	3	2.12 %	
Married	factor	3	0.49 %	
Dependents	factor	5	2.44 %	
Education	factor	2	0.00 %	
Self_Employed	factor	3	5.21 %	
ApplicantIncome	integer	505	0.00 %	×
CoapplicantIncome	numeric	287	0.00 %	×
LoanAmount	integer	204	3.58 %	×
Loan_Amount_Term	integer	11	2.28 %	×
Credit_History	integer	3	8.14 %	
Property_Area	factor	3	0.00 %	
Loan_Status	factor	2	0.00 %	

Variable list

Gender

Feature	Result
Variable type	factor
Number of missing obs.	13 (2.12 %)
Number of unique values	2
Mode	"Male"
Reference category	Female
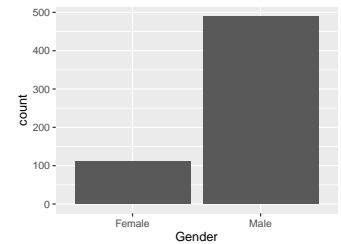
Married

Feature	Result
Variable type	factor
Number of missing obs.	3 (0.49 %)
Number of unique values	2
Mode	"Yes"
Reference category	No
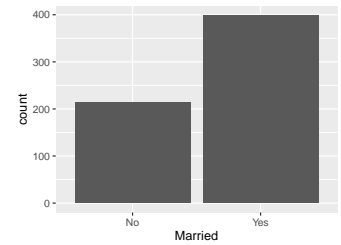
Dependents

Feature	Result
Variable type	factor
Number of missing obs.	15 (2.44 %)
Number of unique values	4
Mode	"0"
Reference category	0
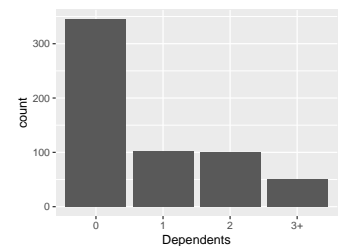
Education

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"Graduate"
Reference category	Graduate
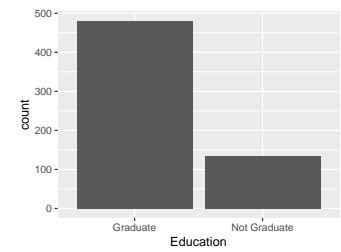
Self_Employed

Feature	Result
Variable type	factor
Number of missing obs.	32 (5.21 %)
Number of unique values	2
Mode	"No"
Reference category	No
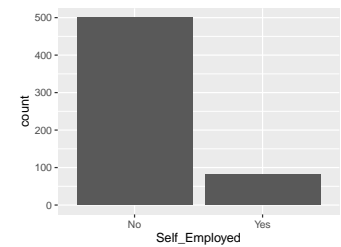
ApplicantIncome

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	505
Median	3812.5
1st and 3rd quartiles	2877.5; 5795
Min. and max.	150; 81000
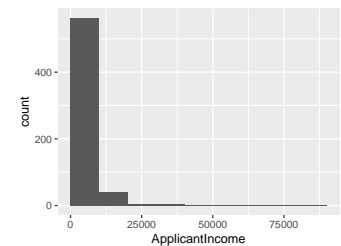
- Note that the following possible outlier values were detected: "150", "210", "416", "645", "674", ..., "39147", "39999", "51763", "63337", "81000" (35 values omitted).

CoapplicantIncome

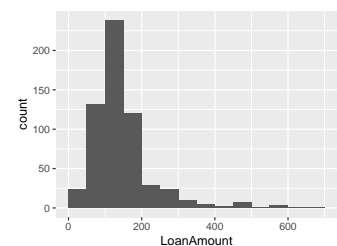| Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	287
Median	1188.5
1st and 3rd quartiles	0; 2297.25
Min. and max.	0; 41667
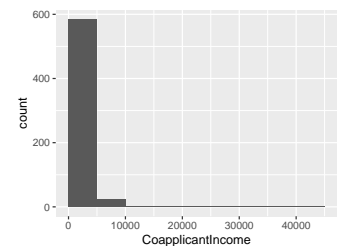
- Note that the following possible outlier values were detected: "6250", "6666", "6667", "7101", "7166", . . . , "10968", "11300", "20000", "33837", "41667" (7 values omitted).

LoanAmount

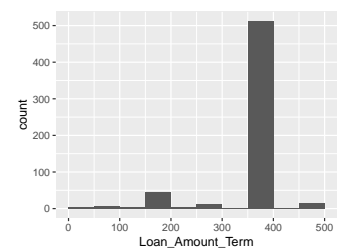| Feature	Result
Variable type	integer
Number of missing obs.	22 (3.58 %)
Number of unique values	203
Median	128
1st and 3rd quartiles	100; 168
Min. and max.	9; 700

- Note that the following possible outlier values were detected: "9", "17", "25", "26", "30", . . . , "500", "570", "600", "650", "700" (24 values omitted).

Loan_Amount_Term

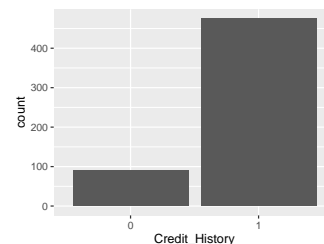| Feature	Result
Variable type	integer
Number of missing obs.	14 (2.28 %)
Number of unique values	10
Median	360
1st and 3rd quartiles	360; 360
Min. and max.	12; 480

- Note that the following possible outlier values were detected: "12", "36", "60", "84", "120", "180", "240", "300", "480".

Credit_History

- 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.	50 (8.14 %)
Number of unique values	2
Mode	"1"
Reference category	0

Property_Area

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	"Semiurban"
Reference category	Rural
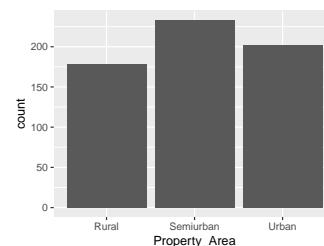
Loan_Status

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"Y"
Reference category	N
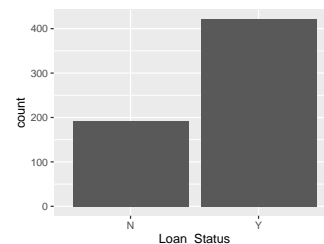
Report generation information:

- Created by: Patel Vinayak (username: `patel`).

- Report creation time: Sun Oct 24 2021 15:53:13

- Report was run from directory: `C:/Users/patel/Documents/Data622`

- dataMaid v1.4.1 [Pkg: 2021-10-08 from CRAN (R 4.1.1)]

- R version 4.1.1 (2021-08-10).

- Platform: x86_64-w64-mingw32/x64 (64-bit)(Windows 10 x64 (build 19043)).

- Function call: `makeDataReport(data = Loan_approval, file = "DataExploratoryReport",     replace = TRUE)`