

# Codebook for train

Autogenerated data summary from dataMaid

2020-08-26 20:56:32

## Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	5778
Number of variables	35

## Codebook summary table

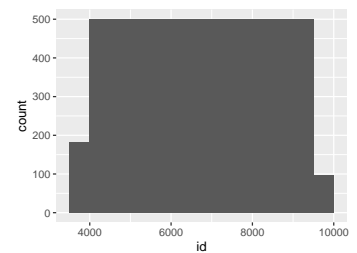
Label	Variable	Class	# unique values	Missing	Description
	id	numeric	5778	0.00 %	
	money_made_inv	numeric	5742	0.00 %	
	acc_now_delinq	character	2	0.00 %	
	acc_open_past_24mths	numeric	27	0.00 %	
	addr_state	factor	8	0.00 %	
	annual_inc	numeric	800	0.00 %	
	application_type	factor	2	0.00 %	
	avg_cur_bal	numeric	2318	0.00 %	
	bc_util	numeric	1034	0.00 %	
	delinq_2yrs	ordered	4	0.00 %	
	dti	numeric	2699	0.00 %	
	emp_length	factor	5	0.00 %	
	grade	factor	7	0.00 %	
	home_ownership	factor	4	0.00 %	
	initial_list_status	factor	2	0.00 %	
	int_rate	numeric	297	0.00 %	
	loan_amnt	numeric	559	0.00 %	
	mort_acc	factor	7	0.00 %	
	num_sats	numeric	45	0.00 %	
	num_tl_120dpd_2m	factor	2	0.00 %	
	num_tl_90g_dpd_24m	factor	2	0.00 %	
	num_tl_30dpd	factor	2	0.00 %	
	out_prncp_inv	numeric	2380	0.00 %	
	pub_rec	factor	3	0.00 %	
	pub_rec_bankruptcies	factor	3	0.00 %	
	purpose	factor	5	0.00 %	
	term	factor	2	0.00 %	
	tot_coll_amt	numeric	660	0.00 %	
	tot_cur_bal	numeric	5711	0.00 %	

Label	Variable	Class	# unique values	Missing	Description
	<b>verification_status</b>	factor	3	0.00 %	
	<b>delinq</b>	factor	2	0.00 %	
	<b>day_from_opening</b>	numeric	497	0.00 %	
	<b>last_credit_pull_from_now</b>	numeric	57	0.00 %	
	<b>day_diff</b>	numeric	1087	0.00 %	
	<b>late</b>	factor	2	0.00 %	

## Variable list

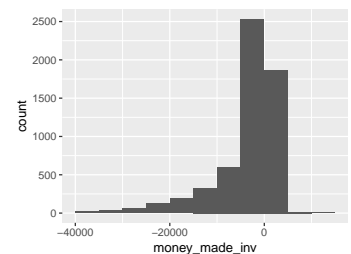
### id

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	5778
Median	6707.5
1st and 3rd quartiles	5263.25; 8151.75
Min. and max.	3819; 9596



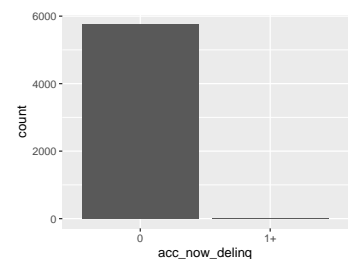
### money\_made\_inv

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	5742
Median	-4.11
1st and 3rd quartiles	-4436.01; 1.87
Min. and max.	-38738.87; 13197.78



### acc\_now\_delinq

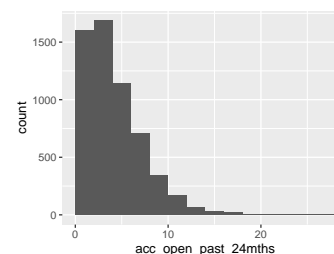
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"



- Observed factor levels: "0", "1+".

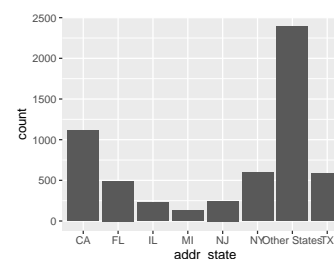
## acc\_open\_past\_24mths

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	27
Median	4
1st and 3rd quartiles	2; 6
Min. and max.	0; 28



## addr\_state

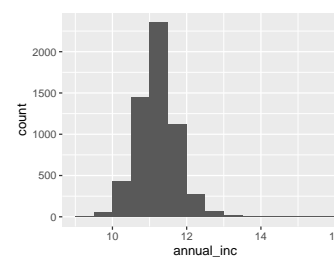
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	8
Mode	"Other States"
Reference category	CA



- Observed factor levels: "CA", "FL", "IL", "MI", "NJ", "NY", "Other States", "TX".

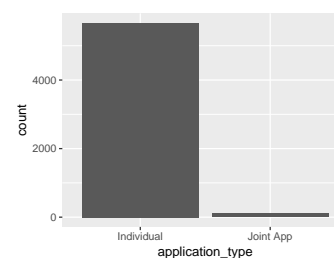
## annual\_inc

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	800
Median	11.16
1st and 3rd quartiles	10.84; 11.51
Min. and max.	9.21; 15.64



## application\_type

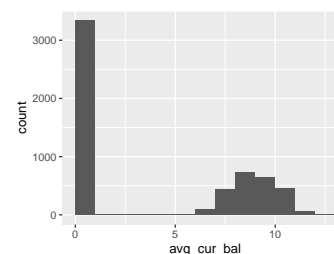
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"Individual"
Reference category	Individual



- Observed factor levels: "Individual", "Joint App".

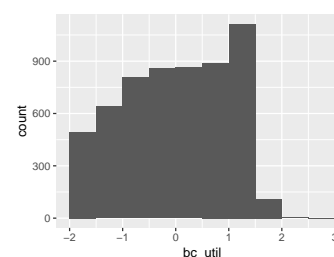
## avg\_cur\_bal

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	2318
Median	0
1st and 3rd quartiles	0; 8.58
Min. and max.	0; 12.43



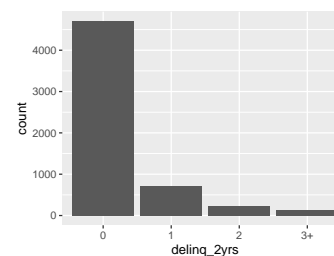
## bc\_util

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1034
Median	0.05
1st and 3rd quartiles	-0.81; 0.88
Min. and max.	-1.98; 2.94



## delinq\_2yrs

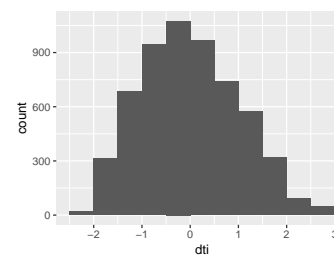
Feature	Result
Variable type	ordered
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"0"
Reference category	0



- Observed factor levels: "0", "1", "2", "3+".

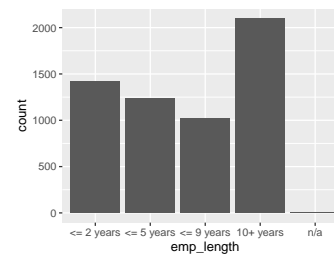
## dti

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	2699
Median	-0.07
1st and 3rd quartiles	-0.75; 0.72
Min. and max.	-2.1; 2.6



## emp\_length

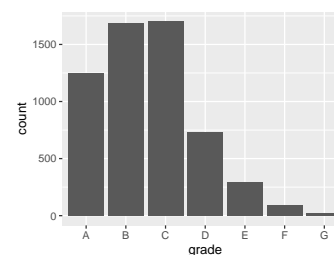
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	5
Mode	"10+ years"
Reference category	<= 2 years



- Observed factor levels: "<= 2 years", "<= 5 years", "<= 9 years", "10+ years", "n/a".

## grade

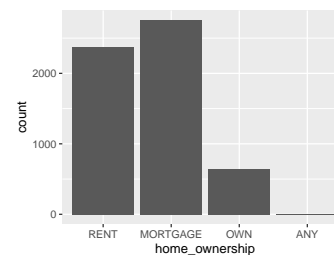
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	7
Mode	"C"
Reference category	A



- Observed factor levels: "A", "B", "C", "D", "E", "F", "G".

## home\_ownership

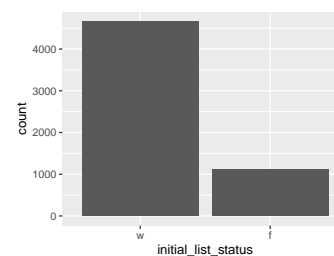
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"MORTGAGE"
Reference category	RENT



- Observed factor levels: "ANY", "MORTGAGE", "OWN", "RENT".

## initial\_list\_status

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

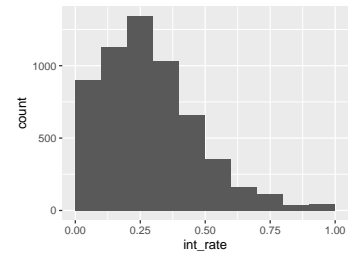


- Observed factor levels: "f", "w".

---

## int\_rate

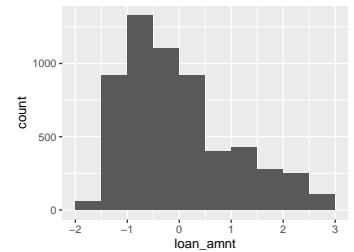
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	297
Median	0.26
1st and 3rd quartiles	0.15; 0.38
Min. and max.	0; 1



---

## loan\_amnt

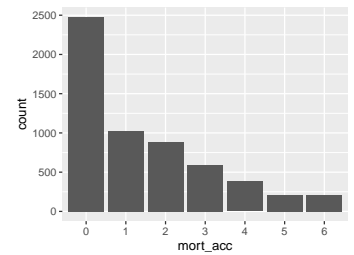
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	559
Median	-0.16
1st and 3rd quartiles	-0.8; 0.58
Min. and max.	-1.56; 2.66



---

## mort\_acc

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	7
Mode	"0"
Reference category	0

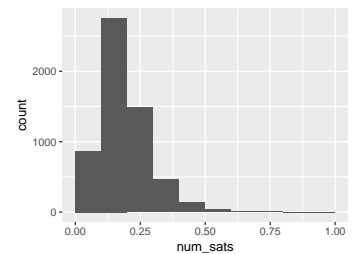


- Observed factor levels: "0", "1", "2", "3", "4", "5", "6".

---

## num\_sats

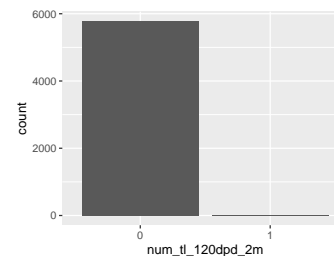
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	45
Median	0.17
1st and 3rd quartiles	0.12; 0.24
Min. and max.	0; 1



## num\_tl\_120dpd\_2m

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

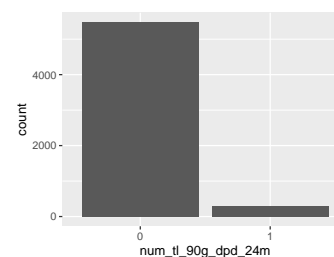
- Observed factor levels: "0", "1".



## num\_tl\_90g\_dpd\_24m

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

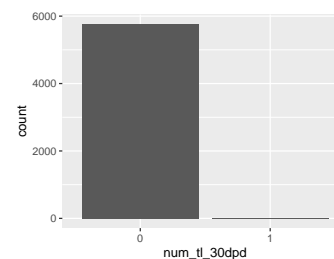
- Observed factor levels: "0", "1".



## num\_tl\_30dpd

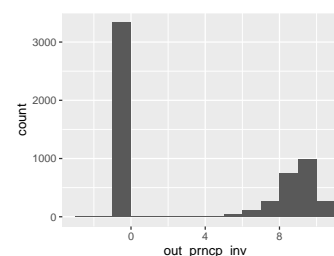
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0

- Observed factor levels: "0", "1".



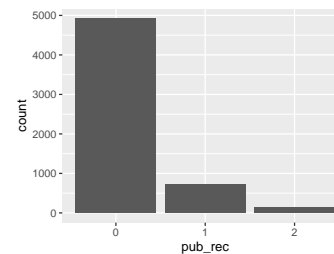
## out\_prncp\_inv

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	2380
Median	0
1st and 3rd quartiles	0; 8.8
Min. and max.	-2.21; 10.57



## pub\_rec

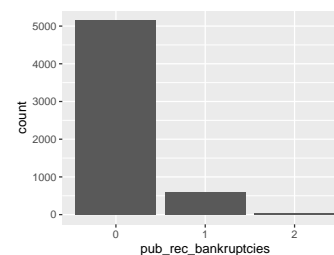
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	"0"
Reference category	0



- Observed factor levels: "0", "1", "2".

## pub\_rec\_bankruptcies

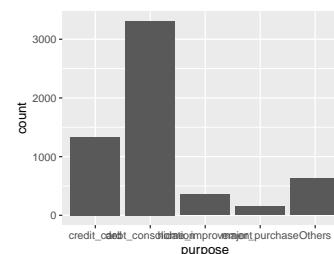
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	"0"
Reference category	0



- Observed factor levels: "0", "1", "2".

## purpose

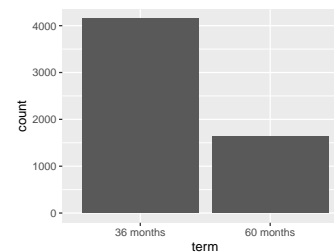
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	5
Mode	"debt_consolidation"
Reference category	credit_card



- Observed factor levels: "credit\_card", "debt\_consolidation", "home\_improvement", "major\_purchase", "Others".

## term

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"36 months"
Reference category	36 months

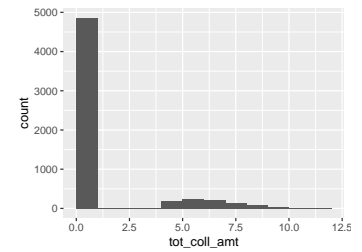


- Observed factor levels: "36 months", "60 months".



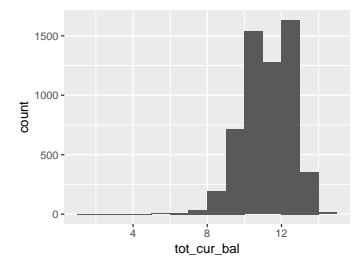
## **tot\_coll\_amt**

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	660
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 11.47



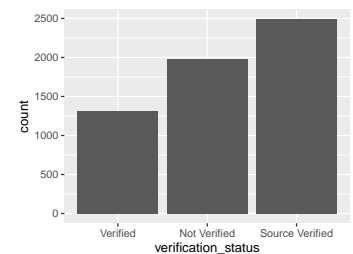
## **tot\_cur\_bal**

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	5711
Median	11.3
1st and 3rd quartiles	10.34; 12.32
Min. and max.	1.1; 14.85



## **verification\_status**

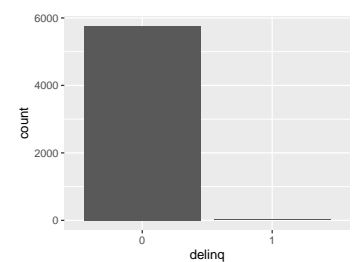
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	"Source Verified"
Reference category	Verified



- Observed factor levels: "Not Verified", "Source Verified", "Verified".

## **delinq**

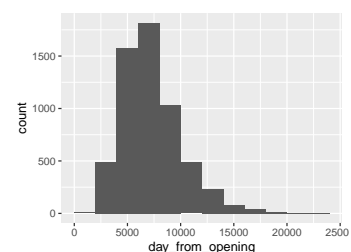
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



- Observed factor levels: "0", "1".

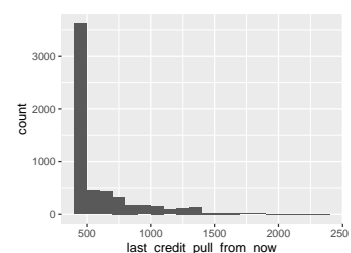
## day\_from\_opening

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	497
Median	6818
1st and 3rd quartiles	5387; 8797
Min. and max.	1766; 22067



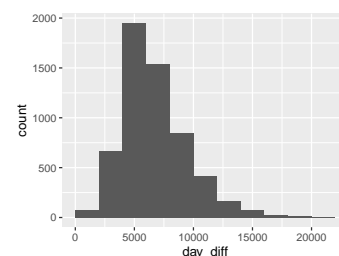
## last\_credit\_pull\_from\_now

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	57
Median	486
1st and 3rd quartiles	486; 639
Min. and max.	486; 2312



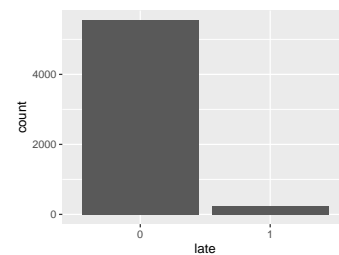
## day\_diff

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1087
Median	6209
1st and 3rd quartiles	4809; 8157.5
Min. and max.	1280; 21581



## late

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



- Observed factor levels: "0", "1".

Report generation information:

- Created by: Junhua Tan (username: JunhuaTan).

- Report creation time: Wed Aug 26 2020 20:56:34
- Report was run from directory: /Users/JunhuaTan/Documents/GitHub/STAT301-3-regression
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 3.6.0)]
- R version 3.6.3 (2020-02-29).
- Platform: x86\_64-apple-darwin15.6.0 (64-bit)(macOS Catalina 10.15.6).
- Function call: `dataMaid::makeDataReport(data = train, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook_train.Rmd", checks = list(character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for train")`