

Overview

Dataset statistics

Number of variables	75
Number of observations	466285
Missing cells	9776224
Missing cells (%)	28.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	266.8 MiB
Average record size in memory	600.0 B

Variable types

Numeric	35
Categorical	22
Boolean	1
Unsupported	17

Alerts

policy_code has constant value "1"	Constant
application_type has constant value "INDIVIDUAL"	Constant
emp_title has a high cardinality: 205475 distinct values	High cardinality
issue_d has a high cardinality: 91 distinct values	High cardinality
url has a high cardinality: 466285 distinct values	High cardinality
desc has a high cardinality: 124436 distinct values	High cardinality
title has a high cardinality: 63099 distinct values	High cardinality
zip_code has a high cardinality: 888 distinct values	High cardinality
earliest_cr_line has a high cardinality: 664 distinct values	High cardinality
last_pymnt_d has a high cardinality: 98 distinct values	High cardinality
next_pymnt_d has a high cardinality: 100 distinct values	High cardinality
last_credit_pull_d has a high cardinality: 103 distinct values	High cardinality
Unnamed: 0 is highly correlated with id and 1 other fields (id, member_id)	High correlation
id is highly correlated with Unnamed: 0 and 3 other fields (Unnamed: 0, member_id, out_prncp, out_prncp_inv)	High correlation
member_id is highly correlated with Unnamed: 0 and 3 other fields (Unnamed: 0, id, out_prncp, out_prncp_inv)	High correlation
loan_amnt is highly correlated with funded_amnt and 7 other fields (funded_amnt, funded_amnt_inv, installment, revol_bal, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)	High correlation
funded_amnt is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, revol_bal, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)	High correlation
funded_amnt_inv is highly correlated with loan_amnt and 6 other fields (loan_amnt, funded_amnt, installment, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)	High correlation
installment is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
annual_inc is highly correlated with tot_cur_bal	High correlation
delinq_2yrs is highly correlated with mths_since_last_delinq and 1 other fields (mths_since_last_delinq, mths_since_last_major_derog)	High correlation
mths_since_last_delinq is highly correlated with delinq_2yrs and 1 other fields (delinq_2yrs, mths_since_last_major_derog)	High correlation
open_acc is highly correlated with total_acc	High correlation
revol_bal is highly correlated with loan_amnt and 2 other fields (loan_amnt, funded_amnt, total_rev_hi_lim)	High correlation
total_acc is highly correlated with open_acc	High correlation
out_prncp is highly correlated with id and 2 other fields (id, member_id, out_prncp_inv)	High correlation
out_prncp_inv is highly correlated with id and 2 other fields (id, member_id, out_prncp)	High correlation
total_pymnt is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
total_pymnt_inv is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
total_rec_prncp is highly correlated with loan_amnt and 6 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, last_pymnt_amnt)	High correlation
total_rec_int is highly correlated with loan_amnt and 5 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv)	High correlation
recoveries is highly correlated with collection_recovery_fee	High correlation
collection_recovery_fee is highly correlated with recoveries	High correlation
last_pymnt_amnt is highly correlated with installment and 3 other fields (installment, total_pymnt, total_pymnt_inv, total_rec_prncp)	High correlation
mths_since_last_major_derog is highly correlated with delinq_2yrs and 1 other fields (delinq_2yrs, mths_since_last_delinq)	High correlation
tot_cur_bal is highly correlated with annual_inc	High correlation
total_rev_hi_lim is highly correlated with revol_bal	High correlation
Unnamed: 0 is highly correlated with member_id	High correlation
id is highly correlated with member_id	High correlation
member_id is highly correlated with Unnamed: 0 and 1 other fields (Unnamed: 0, id)	High correlation

loan_amnt is highly correlated with funded_amnt and 8 other fields (funded_amnt funded_amnt_inv installment_out_prncp_out_prncp_inv total_pymnt total_pymnt_inv total_rec_prncp_total_rec_int)	High correlation
funded_amnt is highly correlated with loan_amnt and 8 other fields (loan_amnt funded_amnt_inv installment_out_prncp_out_prncp_inv total_pymnt total_pymnt_inv total_rec_prncp_total_rec_int)	High correlation
funded_amnt_inv is highly correlated with loan_amnt and 8 other fields (loan_amnt funded_amnt_inv installment_out_prncp_out_prncp_inv total_pymnt total_pymnt_inv total_rec_prncp_total_rec_int)	High correlation
installment is highly correlated with loan_amnt and 6 other fields (loan_amnt funded_amnt funded_amnt_inv total_pymnt total_pymnt_inv total_rec_prncp_total_rec_int)	High correlation
delinq_2yrs is highly correlated with mths_since_last_delinq	High correlation
mths_since_last_delinq is highly correlated with delinq_2yrs and 1 other fields (delinq_2yrs mths_since_last_major_derog)	High correlation
open_acc is highly correlated with total_acc	High correlation
revol_bal is highly correlated with total_rev_hi_lim	High correlation
total_acc is highly correlated with open_acc	High correlation
out_prncp is highly correlated with loan_amnt and 3 other fields (loan_amnt funded_amnt funded_amnt_inv out_prncp_inv)	High correlation
out_prncp_inv is highly correlated with loan_amnt and 3 other fields (loan_amnt funded_amnt funded_amnt_inv out_prncp)	High correlation
total_pymnt is highly correlated with loan_amnt and 7 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt_inv total_rec_prncp_total_rec_int last_pymnt_amnt)	High correlation
total_pymnt_inv is highly correlated with loan_amnt and 7 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt_inv total_rec_prncp_total_rec_int last_pymnt_amnt)	High correlation
total_rec_prncp is highly correlated with loan_amnt and 6 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt_inv total_pymnt_inv last_pymnt_amnt)	High correlation
total_rec_int is highly correlated with loan_amnt and 5 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt_inv total_pymnt_inv)	High correlation
recoveries is highly correlated with collection_recovery_fee	High correlation
collection_recovery_fee is highly correlated with recoveries	High correlation
last_pymnt_amnt is highly correlated with total_pymnt and 2 other fields (total_pymnt total_pymnt_inv total_rec_prncp)	High correlation
mths_since_last_major_derog is highly correlated with mths_since_last_delinq	High correlation
total_rev_hi_lim is highly correlated with revol_bal	High correlation
id is highly correlated with member_id	High correlation
member_id is highly correlated with id	High correlation
loan_amnt is highly correlated with funded_amnt and 5 other fields (funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt total_pymnt_inv total_rec_int)	High correlation
funded_amnt is highly correlated with loan_amnt and 5 other fields (loan_amnt funded_amnt_inv installment_out_pymnt total_pymnt total_pymnt_inv total_rec_int)	High correlation
funded_amnt_inv is highly correlated with loan_amnt and 5 other fields (loan_amnt funded_amnt_inv installment_out_pymnt total_pymnt total_pymnt_inv total_rec_int)	High correlation
installment is highly correlated with loan_amnt and 5 other fields (loan_amnt funded_amnt funded_amnt_inv total_pymnt total_pymnt_inv total_rec_int)	High correlation
delinq_2yrs is highly correlated with mths_since_last_delinq	High correlation
mths_since_last_delinq is highly correlated with delinq_2yrs and 1 other fields (delinq_2yrs mths_since_last_major_derog)	High correlation
open_acc is highly correlated with total_acc	High correlation
revol_bal is highly correlated with total_rev_hi_lim	High correlation
total_acc is highly correlated with open_acc	High correlation
out_prncp is highly correlated with out_prncp_inv	High correlation
out_prncp_inv is highly correlated with out_prncp	High correlation
total_pymnt is highly correlated with loan_amnt and 6 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt total_pymnt_inv total_rec_prncp last_pymnt_amnt)	High correlation
total_pymnt_inv is highly correlated with loan_amnt and 6 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt total_pymnt total_rec_prncp last_pymnt_amnt)	High correlation
total_rec_prncp is highly correlated with total_pymnt and 2 other fields (total_pymnt total_pymnt_inv last_pymnt_amnt)	High correlation
total_rec_int is highly correlated with loan_amnt and 3 other fields (loan_amnt funded_amnt funded_amnt_inv installment_out_pymnt)	High correlation
recoveries is highly correlated with collection_recovery_fee	High correlation

collection_recovery_fee is highly correlated with recoveries	High correlation
last_pymnt_amnt is highly correlated with total_pymnt and 2 other fields (total_pymnt, total_pymnt_inv, total_rec_prncp)	High correlation
mths_since_last_major_derog is highly correlated with mths_since_last_delinq	High correlation
total_rev_hi_lim is highly correlated with revol_bal	High correlation
grade is highly correlated with policy_code and 2 other fields (policy_code, application_type, sub_grade)	High correlation
purpose is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
initial_list_status is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
addr_state is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
last_pymnt_d is highly correlated with policy_code and 2 other fields (policy_code, application_type, next_pymnt_d)	High correlation
issue_d is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
loan_status is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
policy_code is highly correlated with grade and 14 other fields (grade, purpose, initial_list_status, addr_state, last_pymnt_d, issue_d, loan_status, application_type, sub_grade, term, home_ownership, emp_length, pymnt_plan, verification_status, next_pymnt_d)	High correlation
application_type is highly correlated with grade and 14 other fields (grade, purpose, initial_list_status, addr_state, last_pymnt_d, issue_d, loan_status, policy_code, sub_grade, term, home_ownership, emp_length, pymnt_plan, verification_status, next_pymnt_d)	High correlation
sub_grade is highly correlated with grade and 2 other fields (grade, policy_code, application_type)	High correlation
term is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
home_ownership is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
emp_length is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
pymnt_plan is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
verification_status is highly correlated with policy_code and 1 other fields (policy_code, application_type)	High correlation
next_pymnt_d is highly correlated with last_pymnt_d and 2 other fields (last_pymnt_d, policy_code, application_type)	High correlation
Unnamed: 0 is highly correlated with id and 8 other fields (id, member_id, issue_d, mths_since_last_record, initial_list_status, out_prncp_out_prncp_inv, last_pymnt_d, next_pymnt_d)	High correlation
id is highly correlated with Unnamed: 0 and 6 other fields (Unnamed: 0, member_id, issue_d, initial_list_status, out_prncp_out_prncp_inv, last_pymnt_d)	High correlation
member_id is highly correlated with Unnamed: 0 and 6 other fields (Unnamed: 0, id, issue_d, initial_list_status, out_prncp_out_prncp_inv, last_pymnt_d)	High correlation
loan_amnt is highly correlated with funded_amnt and 10 other fields (funded_amnt, funded_amnt_inv, term, installment, out_prncp_out_prncp_inv, total_pymnt_total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
funded_amnt is highly correlated with loan_amnt and 10 other fields (loan_amnt, funded_amnt, funded_amnt_inv, term, installment, out_prncp_out_prncp_inv, total_pymnt_total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
funded_amnt_inv is highly correlated with loan_amnt and 10 other fields (loan_amnt, funded_amnt, term, installment, out_prncp_out_prncp_inv, total_pymnt_total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
term is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, int_rate, sub_grade, out_prncp_out_prncp_inv, total_rec_int)	High correlation
int_rate is highly correlated with term and 3 other fields (term, grade, sub_grade, total_rec_int)	High correlation
installment is highly correlated with loan_amnt and 8 other fields (loan_amnt, funded_amnt, funded_amnt_inv, out_prncp_out_prncp_inv, total_pymnt_total_pymnt_inv, total_rec_prncp, total_rec_int)	High correlation
grade is highly correlated with int_rate and 1 other fields (int_rate, sub_grade)	High correlation
sub_grade is highly correlated with term and 2 other fields (term, int_rate, grade)	High correlation
issue_d is highly correlated with Unnamed: 0 and 7 other fields (Unnamed: 0, id, member_id, loan_status, mths_since_last_record, initial_list_status, last_pymnt_d, next_pymnt_d)	High correlation
loan_status is highly correlated with issue_d and 4 other fields (issue_d, out_prncp_out_prncp_inv, last_pymnt_d, next_pymnt_d)	High correlation
inq_last_6mths is highly correlated with next_pymnt_d	High correlation
mths_since_last_delinq is highly correlated with mths_since_last_major_derog	High correlation
mths_since_last_record is highly correlated with Unnamed: 0 and 3 other fields (Unnamed: 0, issue_d, last_pymnt_d, next_pymnt_d)	High correlation
open_acc is highly correlated with total_acc	High correlation
revol_bal is highly correlated with tot_cur_bal and 1 other fields (tot_cur_bal, total_rev_hi_lim)	High correlation
total_acc is highly correlated with open_acc	High correlation
initial_list_status is highly correlated with Unnamed: 0 and 3 other fields (Unnamed: 0, id, member_id, issue_d)	High correlation

out_prncp is highly correlated with Unnamed: 0 and 11 other fields (Unnamed: 0, id, member_id, loan_amnt, funded_amnt, funded_amnt_inv, term, installment, loan_status, out_prncp_inv, total_rec_int, last_pymnt_d)	High correlation
out_prncp_inv is highly correlated with Unnamed: 0 and 11 other fields (Unnamed: 0, id, member_id, loan_amnt, funded_amnt, funded_amnt_inv, term, installment, loan_status, out_prncp, total_rec_int, last_pymnt_d)	High correlation
total_pymnt is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
total_pymnt_inv is highly correlated with loan_amnt and 7 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_rec_prncp, total_rec_int, last_pymnt_amnt)	High correlation
total_rec_prncp is highly correlated with loan_amnt and 6 other fields (loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, last_pymnt_amnt)	High correlation
total_rec_int is highly correlated with loan_amnt and 9 other fields (loan_amnt, funded_amnt, funded_amnt_inv, term, int_rate, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv)	High correlation
recoveries is highly correlated with collection_recovery_fee	High correlation
collection_recovery_fee is highly correlated with recoveries	High correlation
last_pymnt_d is highly correlated with Unnamed: 0 and 8 other fields (Unnamed: 0, id, member_id, issue_d, loan_status, mths_since_last_record, out_prncp, out_prncp_inv, next_pymnt_d)	High correlation
last_pymnt_amnt is highly correlated with loan_amnt and 6 other fields (loan_amnt, funded_amnt, funded_amnt_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, next_pymnt_d)	High correlation
next_pymnt_d is highly correlated with Unnamed: 0 and 6 other fields (Unnamed: 0, issue_d, loan_status, ing_last_6mths, mths_since_last_record, last_pymnt_d, last_pymnt_amnt)	High correlation
mths_since_last_major_derog is highly correlated with mths_since_last_delinq	High correlation
tot_cur_bal is highly correlated with revol_bal	High correlation
total_rev_hi_lim is highly correlated with revol_bal	High correlation
emp_title has 27588 (5.9%) missing values	Missing
emp_length has 21008 (4.5%) missing values	Missing
desc has 340302 (73.0%) missing values	Missing
mths_since_last_delinq has 250351 (53.7%) missing values	Missing
mths_since_last_record has 403647 (86.6%) missing values	Missing
next_pymnt_d has 227214 (48.7%) missing values	Missing
mths_since_last_major_derog has 367311 (78.8%) missing values	Missing
annual_inc_joint has 466285 (100.0%) missing values	Missing
dti_joint has 466285 (100.0%) missing values	Missing
verification_status_joint has 466285 (100.0%) missing values	Missing
tot_coll_amt has 70276 (15.1%) missing values	Missing
tot_cur_bal has 70276 (15.1%) missing values	Missing
open_acc_6m has 466285 (100.0%) missing values	Missing
open_il_6m has 466285 (100.0%) missing values	Missing
open_il_12m has 466285 (100.0%) missing values	Missing
open_il_24m has 466285 (100.0%) missing values	Missing
mths_since_rcnt_il has 466285 (100.0%) missing values	Missing
total_bal_il has 466285 (100.0%) missing values	Missing
il_util has 466285 (100.0%) missing values	Missing
open_rv_12m has 466285 (100.0%) missing values	Missing
open_rv_24m has 466285 (100.0%) missing values	Missing
max_bal_bc has 466285 (100.0%) missing values	Missing
all_util has 466285 (100.0%) missing values	Missing
total_rev_hi_lim has 70276 (15.1%) missing values	Missing
inq_fi has 466285 (100.0%) missing values	Missing
total_cu_tl has 466285 (100.0%) missing values	Missing
inq_last_12m has 466285 (100.0%) missing values	Missing

annual_inc is highly skewed ($\gamma_1 = 29.56890538$)	Skewed
collection_recovery_fee is highly skewed ($\gamma_1 = 22.23609553$)	Skewed
collections_12_mths_ex_med is highly skewed ($\gamma_1 = 31.22030218$)	Skewed
acc_now_delinq is highly skewed ($\gamma_1 = 20.79829116$)	Skewed
tot_coll_amt is highly skewed ($\gamma_1 = 618.2730112$)	Skewed
total_rev_hi_lim is highly skewed ($\gamma_1 = 100.5855901$)	Skewed
Unnamed: 0 is uniformly distributed	Uniform
url is uniformly distributed	Uniform
desc is uniformly distributed	Uniform
Unnamed: 0 has unique values	Unique
id has unique values	Unique
member_id has unique values	Unique
url has unique values	Unique
annual_inc_joint is an unsupported type, check if it needs cleaning or further analysis	Unsupported
dti_joint is an unsupported type, check if it needs cleaning or further analysis	Unsupported
verification_status_joint is an unsupported type, check if it needs cleaning or further analysis	Unsupported
open_acc_6m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
open_il_6m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
open_il_12m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
open_il_24m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
mths_since_rcnt_il is an unsupported type, check if it needs cleaning or further analysis	Unsupported
total_bal_il is an unsupported type, check if it needs cleaning or further analysis	Unsupported
il_util is an unsupported type, check if it needs cleaning or further analysis	Unsupported
open_rv_12m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
open_rv_24m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
max_bal_bc is an unsupported type, check if it needs cleaning or further analysis	Unsupported
all_util is an unsupported type, check if it needs cleaning or further analysis	Unsupported
inq_fi is an unsupported type, check if it needs cleaning or further analysis	Unsupported
total_cu_tl is an unsupported type, check if it needs cleaning or further analysis	Unsupported
inq_last_12m is an unsupported type, check if it needs cleaning or further analysis	Unsupported
delinq_2yrs has 382954 (82.1%) zeros	Zeros
inq_last_6mths has 241498 (51.8%) zeros	Zeros
pub_rec has 404893 (86.8%) zeros	Zeros
out_prncp has 229999 (49.3%) zeros	Zeros
out_prncp_inv has 229999 (49.3%) zeros	Zeros
total_rec_late_fee has 455607 (97.7%) zeros	Zeros
recoveries has 441914 (94.8%) zeros	Zeros
collection_recovery_fee has 443079 (95.0%) zeros	Zeros
collections_12_mths_ex_med has 462226 (99.1%) zeros	Zeros
acc_now_delinq has 464524 (99.6%) zeros	Zeros
tot_coll_amt has 346259 (74.3%) zeros	Zeros

Reproduction

Analysis started	2021-12-30 20:13:49.227944
Analysis finished	2021-12-30 20:34:24.346811
Duration	20 minutes and 35.12 seconds
Software version	pandas-profiling v3.1.0 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22EDA%20using%20Pandas%20Profiling%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%8

Variables

Unnamed: 0

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: id, member_id)

HIGH CORRELATION (This variable has a high correlation with 1 fields: member_id)

HIGH CORRELATION (This variable has a high correlation with 9 fields: id, member_id, issue_d, mths_since_last_record, initial_list_status, out_prncp, out_prncp_inv, last_pymnt_d, next_pymnt_d)

UNIFORM

UNIQUE

Distinct	466285
----------	--------

Distinct (%)	100.0%
--------------	--------

Missing	0
---------	---

Missing (%)	0.0%
-------------	------

Infinite	0
----------	---

Infinite (%)	0.0%
--------------	------

Mean	233142
------	--------

Minimum	0
---------	---

Maximum	466284
---------	--------

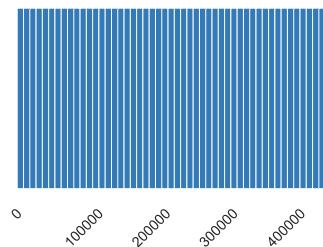
Zeros	1
-------	---

Zeros (%)	< 0.1%
-----------	--------

Negative	0
----------	---

Negative (%)	0.0%
--------------	------

Memory size	3.6 MiB
-------------	---------



Quantile statistics

Minimum	0
---------	---

5-th percentile	23314.2
-----------------	---------

Q1	116571
----	--------

median	233142
--------	--------

Q3	349713
----	--------

95-th percentile	442969.8
------------------	----------

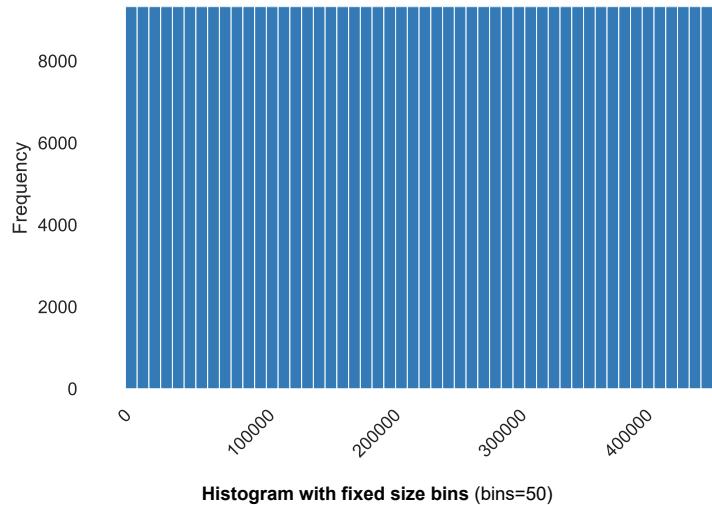
Maximum	466284
---------	--------

Range	466284
-------	--------

Interquartile range (IQR)	233142
---------------------------	--------

Descriptive statistics

Standard deviation	134605.0295
Coefficient of variation (CV)	0.5773521265
Kurtosis	-1.2
Mean	233142
Median Absolute Deviation (MAD)	116571
Skewness	$-1.728895215 \times 10^{-15}$
Sum	$1.087106175 \times 10^{11}$
Variance	$1.811851396 \times 10^{10}$
Monotonicity	Strictly increasing



Value	Count	Frequency (%)
0	1	< 0.1%
310839	1	< 0.1%
310865	1	< 0.1%
310864	1	< 0.1%
310863	1	< 0.1%
310862	1	< 0.1%
310861	1	< 0.1%
310860	1	< 0.1%
310859	1	< 0.1%
310858	1	< 0.1%
Other values (466275)	466275	> 99.9%

Value	Count	Frequency (%)
0	1	< 0.1%
1	1	< 0.1%
2	1	< 0.1%
3	1	< 0.1%
4	1	< 0.1%
5	1	< 0.1%
6	1	< 0.1%
7	1	< 0.1%
8	1	< 0.1%
9	1	< 0.1%

Value	Count	Frequency (%)
466284	1	< 0.1%
466283	1	< 0.1%
466282	1	< 0.1%
466281	1	< 0.1%
466280	1	< 0.1%
466279	1	< 0.1%
466278	1	< 0.1%
466277	1	< 0.1%
466276	1	< 0.1%
466275	1	< 0.1%

id

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 4 fields: Unnamed: 0, member_id, out_prncp, out_prncp_inv)

HIGH CORRELATION (This variable has a high correlation with 1 fields: member_id)

HIGH CORRELATION (This variable has a high correlation with 1 fields: member_id)

HIGH CORRELATION (This variable has a high correlation with 7 fields: Unnamed: 0, member_id, issue_d, initial_list_status, out_prncp, out_prncp_inv, last_pymnt_d)

UNIQUE

Distinct 466285

Distinct (%) 100.0%

Missing 0

Missing (%) 0.0%

Infinite 0

Infinite (%) 0.0%

Mean 13079729.91

Minimum 54734

Maximum 38098114

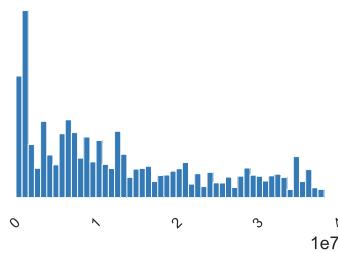
Zeros 0

Zeros (%) 0.0%

Negative 0

Negative (%) 0.0%

Memory size 3.6 MiB



Quantile statistics

Minimum 54734

5-th percentile 676036.4

Q1 3639987

median 10107897

Q3 20731209

95-th percentile 34453013.6

Maximum 38098114

Range 38043380

Interquartile range (IQR) 17091222

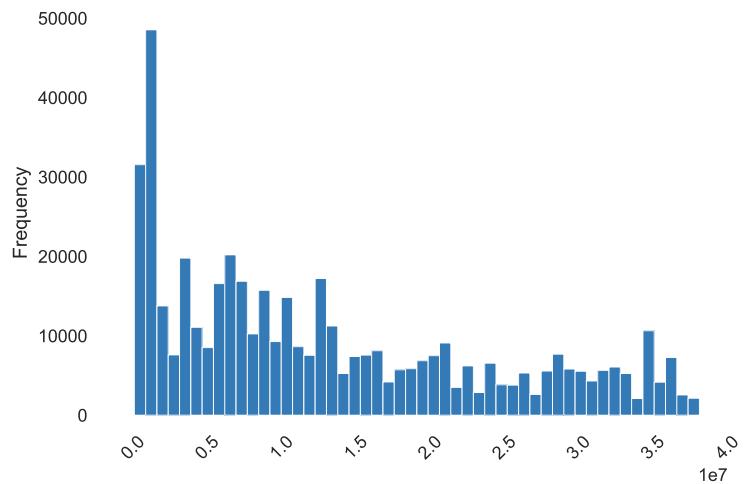
Descriptive statistics

Standard deviation 10893711.62

Coefficient of variation (CV) 0.8328697689

Kurtosis -0.7276589587

Mean	13079729.91
Median Absolute Deviation (MAD)	7801562
Skewness	0.6924558802
Sum	$6.098881859 \times 10^{12}$
Variance	$1.18672953 \times 10^{14}$
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1077501	1	< 0.1%
26180404	1	< 0.1%
26219322	1	< 0.1%
26239577	1	< 0.1%
26290537	1	< 0.1%
26230483	1	< 0.1%
26350584	1	< 0.1%
26269915	1	< 0.1%
26340359	1	< 0.1%
26299926	1	< 0.1%
Other values (466275)	466275	> 99.9%

Value	Count	Frequency (%)
54734	1	< 0.1%
55521	1	< 0.1%
55742	1	< 0.1%
56413	1	< 0.1%
57167	1	< 0.1%
57245	1	< 0.1%
57416	1	< 0.1%
58524	1	< 0.1%
58915	1	< 0.1%
59006	1	< 0.1%

Value	Count	Frequency (%)
38098114	1	< 0.1%
37854444	1	< 0.1%
37854423	1	< 0.1%
37854157	1	< 0.1%
37854137	1	< 0.1%
37842129	1	< 0.1%
37841931	1	< 0.1%
37841841	1	< 0.1%
37841820	1	< 0.1%
37841743	1	< 0.1%

member_id

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 4 fields: Unnamed: 0, id, out_prncp, out_prncp_inv)

HIGH CORRELATION (This variable has a high correlation with 2 fields: Unnamed: 0, id)

HIGH CORRELATION (This variable has a high correlation with 1 fields: id)

HIGH CORRELATION (This variable has a high correlation with 7 fields: Unnamed: 0, id, issue_d, initial_list_status, out_prncp, out_prncp_inv, last_pymnt_d)

UNIQUE

Distinct	466285
----------	--------

Distinct (%)	100.0%
--------------	--------

Missing	0
---------	---

Missing (%)	0.0%
-------------	------

Infinite	0
----------	---

Infinite (%)	0.0%
--------------	------

Mean	14597660.62
------	-------------

Minimum	70473
---------	-------

Maximum	40860827
---------	----------

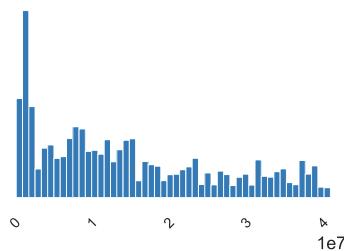
Zeros	0
-------	---

Zeros (%)	0.0%
-----------	------

Negative	0
----------	---

Negative (%)	0.0%
--------------	------

Memory size	3.6 MiB
-------------	---------



Quantile statistics

Minimum	70473
---------	-------

5-th percentile	863451.8
-----------------	----------

Q1	4379705
----	---------

median	11941075
--------	----------

Q3	23001541
----	----------

95-th percentile	37115953.4
------------------	------------

Maximum	40860827
---------	----------

Range	40790354
-------	----------

Interquartile range (IQR)	18621836
---------------------------	----------

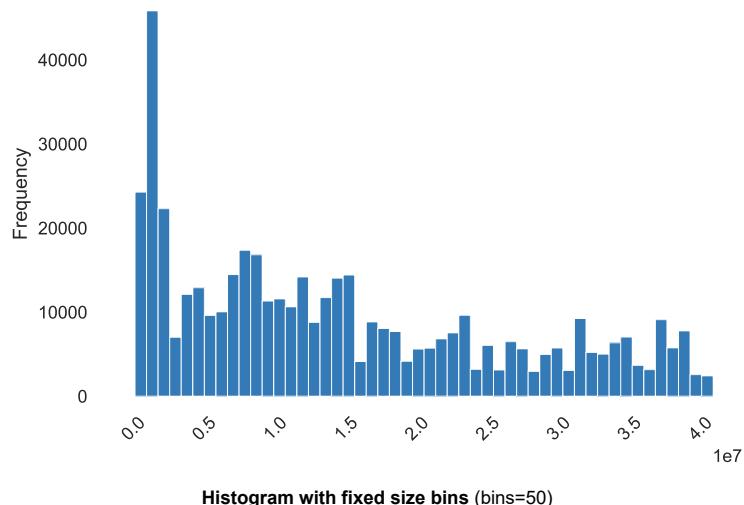
Descriptive statistics

Standard deviation	11682367.44
--------------------	-------------

Coefficient of variation (CV)	0.8002903851
-------------------------------	--------------

Kurtosis	-0.7974834576
----------	---------------

Mean	14597660.62
Median Absolute Deviation (MAD)	9034255
Skewness	0.6213934744
Sum	$6.806670184 \times 10^{12}$
Variance	$1.36477709 \times 10^{14}$
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1296599	1	< 0.1%
28653081	1	< 0.1%
28692177	1	< 0.1%
28702376	1	< 0.1%
28763241	1	< 0.1%
28713125	1	< 0.1%
28823219	1	< 0.1%
28742665	1	< 0.1%
28813062	1	< 0.1%
28772662	1	< 0.1%
Other values (466275)	466275	> 99.9%

Value	Count	Frequency (%)
70473	1	< 0.1%
70626	1	< 0.1%
70651	1	< 0.1%
70681	1	< 0.1%
70694	1	< 0.1%
70699	1	< 0.1%
70735	1	< 0.1%
70868	1	< 0.1%
70879	1	< 0.1%
70978	1	< 0.1%

Value	Count	Frequency (%)
40860827	1	< 0.1%
40617199	1	< 0.1%
40617173	1	< 0.1%
40616878	1	< 0.1%
40616855	1	< 0.1%
40605224	1	< 0.1%
40605012	1	< 0.1%
40604912	1	< 0.1%
40604886	1	< 0.1%
40604792	1	< 0.1%

loan_amnt

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

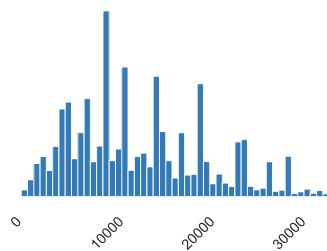
CORRELATION (This variable has a high correlation with 8 fields: funded_amnt, funded_amnt_inv, installment, revol_bal, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 9 fields: funded_amnt, funded_amnt_inv, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 6 fields: funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 11 fields: funded_amnt, funded_amnt_inv, term, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)

Distinct	1352
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	14317.27758
Minimum	500
Maximum	35000
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MB



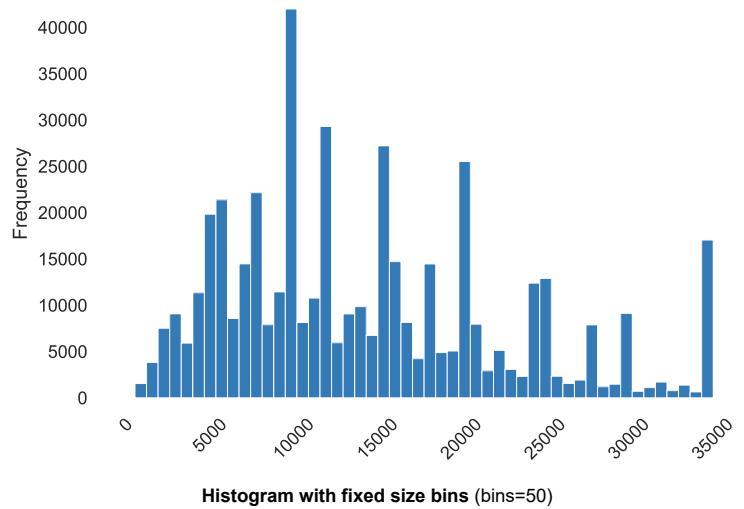
Quantile statistics

Minimum	500
5-th percentile	3500
Q1	8000
median	12000
Q3	20000
95-th percentile	30200
Maximum	35000
Range	34500
Interquartile range (IQR)	12000

Descriptive statistics

Standard deviation	8286.509164
Coefficient of variation (CV)	0.5787768743

Kurtosis	-0.1348784749
Mean	14317.27758
Median Absolute Deviation (MAD)	5500
Skewness	0.7309798832
Sum	6675931775
Variance	68666234.13
Monotonicity	Not monotonic



Value	Count	Frequency (%)
10000	33023	7.1%
12000	25519	5.5%
15000	23486	5.0%
20000	22759	4.9%
35000	16596	3.6%
8000	14867	3.2%
6000	13985	3.0%
5000	13879	3.0%
16000	12069	2.6%
18000	11616	2.5%
Other values (1342)	278486	59.7%

Value	Count	Frequency (%)
500	11	< 0.1%
550	1	< 0.1%
600	6	< 0.1%
700	3	< 0.1%
725	1	< 0.1%
750	1	< 0.1%
800	3	< 0.1%
850	1	< 0.1%
900	4	< 0.1%
925	1	< 0.1%

Value	Count	Frequency (%)
35000	16596	3.6%
34975	15	< 0.1%
34950	2	< 0.1%
34925	2	< 0.1%
34900	5	< 0.1%
34875	42	< 0.1%
34850	5	< 0.1%
34825	1	< 0.1%
34800	32	< 0.1%
34775	1	< 0.1%

funded_amnt

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

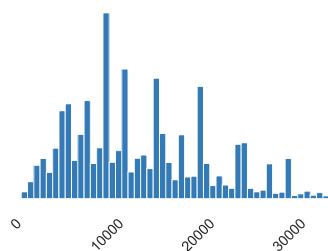
CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt_inv, installment, revol_bal, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 9 fields: loan_amnt, funded_amnt_inv, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 6 fields: loan_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 11 fields: loan_amnt, funded_amnt_inv, term, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)

Distinct	1354
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	14291.80104
Minimum	500
Maximum	35000
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



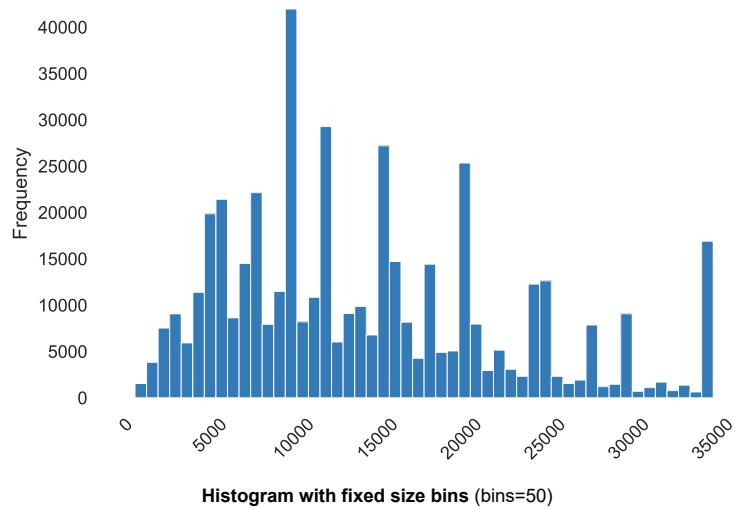
Quantile statistics

Minimum	500
5-th percentile	3500
Q1	8000
median	12000
Q3	20000
95-th percentile	30000
Maximum	35000
Range	34500
Interquartile range (IQR)	12000

Descriptive statistics

Standard deviation	8274.3713
Coefficient of variation (CV)	0.578959312

Kurtosis	-0.1241203906
Mean	14291.80104
Median Absolute Deviation (MAD)	5500
Skewness	0.7348454478
Sum	6664052450
Variance	68465220.41
Monotonicity	Not monotonic



Value	Count	Frequency (%)
10000	32930	7.1%
12000	25426	5.5%
15000	23370	5.0%
20000	22571	4.8%
35000	16458	3.5%
8000	14854	3.2%
6000	13971	3.0%
5000	13865	3.0%
16000	11982	2.6%
18000	11540	2.5%
Other values (1344)	279318	59.9%

Value	Count	Frequency (%)
500	11	< 0.1%
550	1	< 0.1%
600	6	< 0.1%
700	3	< 0.1%
725	1	< 0.1%
750	1	< 0.1%
800	3	< 0.1%
850	1	< 0.1%
900	4	< 0.1%
925	1	< 0.1%

Value	Count	Frequency (%)
35000	16458	3.5%
34975	15	< 0.1%
34950	2	< 0.1%
34925	2	< 0.1%
34900	5	< 0.1%
34875	42	< 0.1%
34850	5	< 0.1%
34825	1	< 0.1%
34800	31	< 0.1%
34775	1	< 0.1%

funded_amnt_inv

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

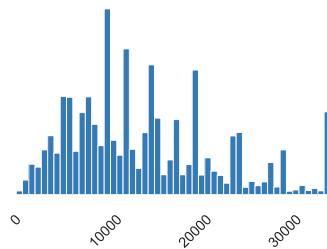
CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, installment, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 9 fields: loan_amnt, funded_amnt, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 6 fields: loan_amnt, funded_amnt, installment, total_pymnt, total_pymnt_inv, total_rec_int)

HIGH CORRELATION (This variable has a high correlation with 11 fields: loan_amnt, funded_amnt, term, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt)

Distinct	9854
Distinct (%)	2.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	14222.32989
Minimum	0
Maximum	35000
Zeros	233
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



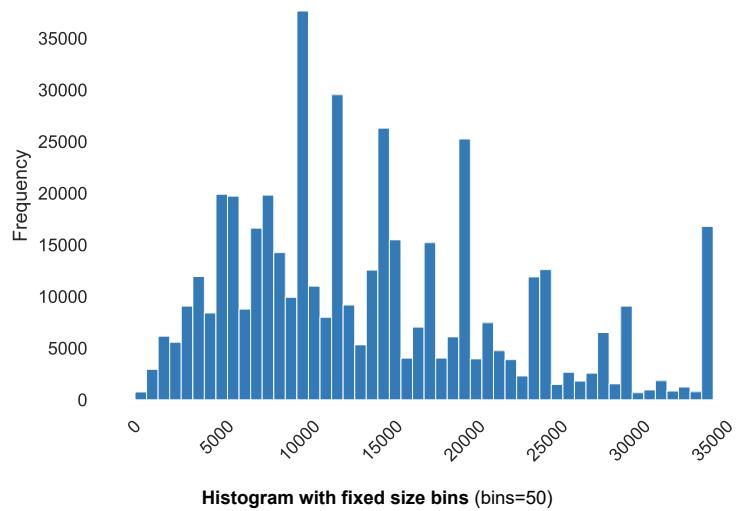
Quantile statistics

Minimum	0
5-th percentile	3250
Q1	8000
median	12000
Q3	19950
95-th percentile	30000
Maximum	35000
Range	35000
Interquartile range (IQR)	11950

Descriptive statistics

Standard deviation	8297.637788
Coefficient of variation (CV)	0.5834232403

Kurtosis	-0.1238605674
Mean	14222.32989
Median Absolute Deviation (MAD)	5500
Skewness	0.7302504405
Sum	6631659092
Variance	68850792.87
Monotonicity	Not monotonic



Value	Count	Frequency (%)
10000	28683	6.2%
12000	21732	4.7%
15000	19473	4.2%
20000	18070	3.9%
8000	13232	2.8%
35000	13130	2.8%
6000	12635	2.7%
5000	12547	2.7%
16000	9906	2.1%
18000	9511	2.0%
Other values (9844)	307366	65.9%

Value	Count	Frequency (%)
0	233	< 0.1%
0.000121098	1	< 0.1%
0.000185369	1	< 0.1%
0.000242056	1	< 0.1%
0.000531133	1	< 0.1%
0.000571783	1	< 0.1%
0.000654607	1	< 0.1%
0.000899697	1	< 0.1%
0.00092203	1	< 0.1%
0.00108361	1	< 0.1%

Value	Count	Frequency (%)
35000	13130	2.8%
34997.35245	1	< 0.1%
34995.37946	1	< 0.1%
34993.65539	1	< 0.1%
34993.32571	1	< 0.1%
34993.26306	1	< 0.1%
34993.19696	1	< 0.1%
34992.98044	1	< 0.1%
34992.76781	1	< 0.1%
34992.48861	1	< 0.1%

term

Categorical

HIGH**CORRELATION** (This variable has a high correlation with 2 fields: policy_code, application_type)**HIGH CORRELATION** (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, int_rate, sub_grade, out_prncp, out_prncp_inv, total_rec_int)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	10
Median length	10
Mean length	10
Min length	10

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

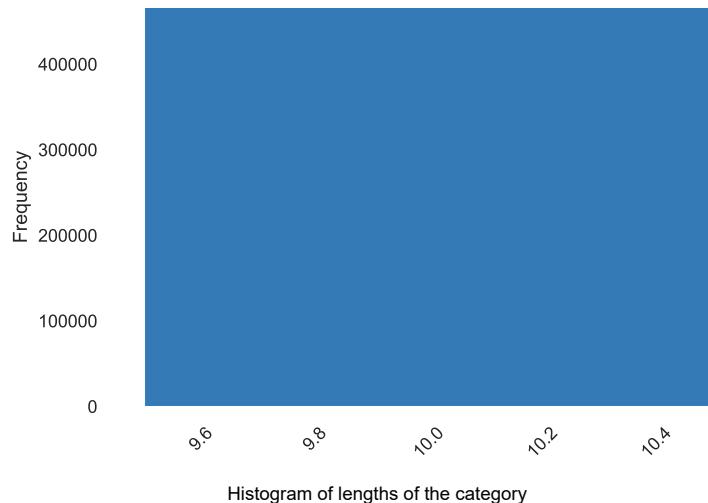
Sample

1st row	36 months
2nd row	60 months
3rd row	36 months
4th row	36 months
5th row	60 months

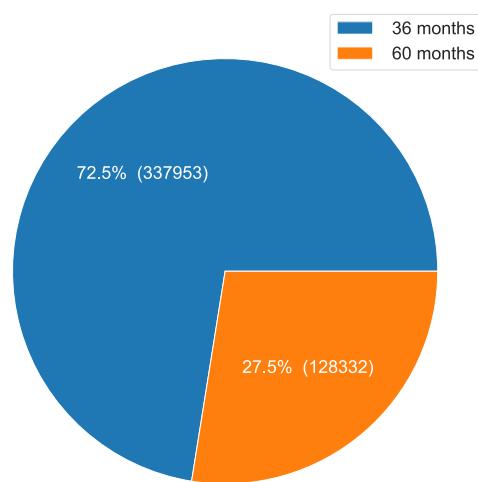
Common Values

Value	Count	Frequency (%)
36 months	337953	72.5%
60 months	128332	27.5%

Length



Pie chart



Value	Count	Frequency (%)
months	466285	50.0%
36	337953	36.2%
60	128332	13.8%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

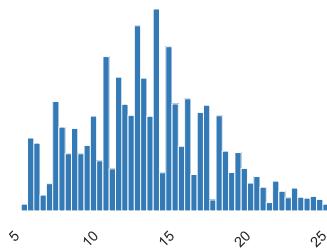
int_rate

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 4 fields: term, grade, sub_grade, total_rec_int)

Distinct	506
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	13.82923594
Minimum	5.42
Maximum	26.06
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



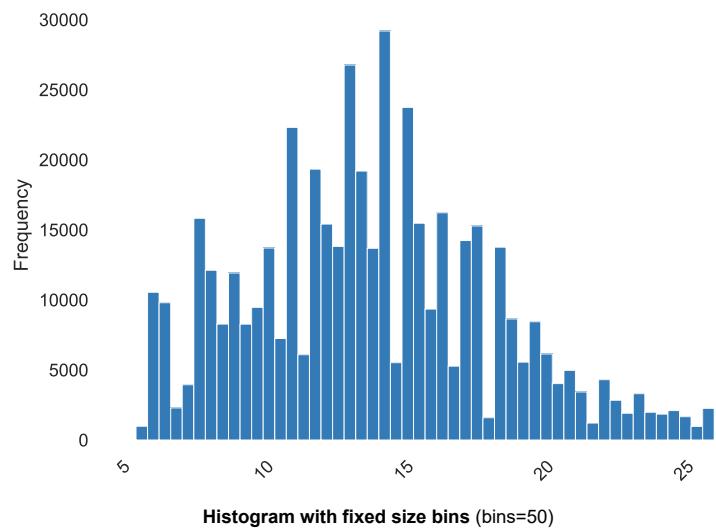
Quantile statistics

Minimum	5.42
5-th percentile	6.99
Q1	10.99
median	13.66
Q3	16.49
95-th percentile	21.7
Maximum	26.06
Range	20.64
Interquartile range (IQR)	5.5

Descriptive statistics

Standard deviation	4.357586583
Coefficient of variation (CV)	0.3150995906
Kurtosis	-0.2625169147
Mean	13.82923594
Median Absolute Deviation (MAD)	2.67
Skewness	0.3349382355

Sum	6448365.28
Variance	18.98856082
Monotonicity	Not monotonic



Value	Count	Frequency (%)
12.99	15846	3.4%
10.99	14837	3.2%
15.61	13074	2.8%
8.9	11424	2.5%
16.29	10587	2.3%
11.99	10114	2.2%
7.9	10079	2.2%
6.03	9846	2.1%
13.98	9803	2.1%
12.49	9727	2.1%
Other values (496)	350948	75.3%

Value	Count	Frequency (%)
5.42	573	0.1%
5.79	410	0.1%
5.99	347	0.1%
6	100	< 0.1%
6.03	9846	2.1%
6.17	252	0.1%
6.39	58	< 0.1%
6.49	4026	0.9%
6.54	307	0.1%
6.62	5409	1.2%

Value	Count	Frequency (%)
26.06	264	0.1%
25.99	277	0.1%
25.89	428	0.1%
25.83	576	0.1%
25.8	715	0.2%
25.57	801	0.2%
25.28	171	< 0.1%
24.99	1202	0.3%
24.89	352	0.1%
24.83	112	< 0.1%

installment

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

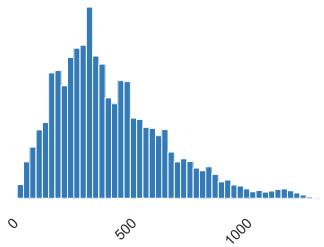
CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt).

HIGH CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int).

HIGH CORRELATION (This variable has a high correlation with 6 fields: loan_amnt, funded_amnt, funded_amnt_inv, total_pymnt, total_pymnt_inv, total_rec_int).

HIGH CORRELATION (This variable has a high correlation with 9 fields: loan_amnt, funded_amnt, funded_amnt_inv, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int).

Distinct	55622
Distinct (%)	11.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	432.0612007
Minimum	15.67
Maximum	1409.99
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



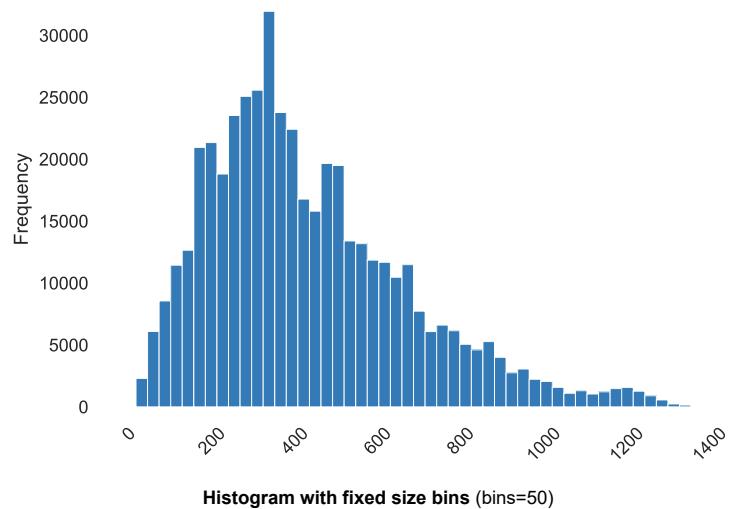
Quantile statistics

Minimum	15.67
5-th percentile	114.51
Q1	256.69
median	379.89
Q3	566.58
95-th percentile	900.93
Maximum	1409.99
Range	1394.32
Interquartile range (IQR)	309.89

Descriptive statistics

Standard deviation	243.4855499
Coefficient of variation (CV)	0.5635441216

Kurtosis	0.7679860196
Mean	432.0612007
Median Absolute Deviation (MAD)	147.05
Skewness	0.949293412
Sum	201463657
Variance	59285.21299
Monotonicity	Not monotonic



Value	Count	Frequency (%)
327.34	1246	0.3%
336.9	1191	0.3%
332.1	978	0.2%
491.01	881	0.2%
392.81	855	0.2%
317.54	848	0.2%
332.72	824	0.2%
334.49	775	0.2%
304.36	775	0.2%
337.47	760	0.2%
Other values (55612)	457152	98.0%

Value	Count	Frequency (%)
15.67	1	< 0.1%
15.69	1	< 0.1%
15.75	1	< 0.1%
15.76	1	< 0.1%
15.91	1	< 0.1%
16.08	1	< 0.1%
16.25	1	< 0.1%
16.31	1	< 0.1%
16.47	1	< 0.1%
16.73	1	< 0.1%

Value	Count	Frequency (%)
1409.99	2	< 0.1%
1408.13	5	< 0.1%
1407.01	7	< 0.1%
1406.45	9	< 0.1%
1404.95	1	< 0.1%
1402.17	4	< 0.1%
1396.79	1	< 0.1%
1391.41	4	< 0.1%
1388.45	2	< 0.1%
1382.36	8	< 0.1%

grade
Categorical

HIGH

CORRELATION (This variable has a high correlation with 3 fields: policy_code, application_type, sub_grade)
HIGH CORRELATION (This variable has a high correlation with 2 fields: int_rate, sub_grade)

Distinct	7
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

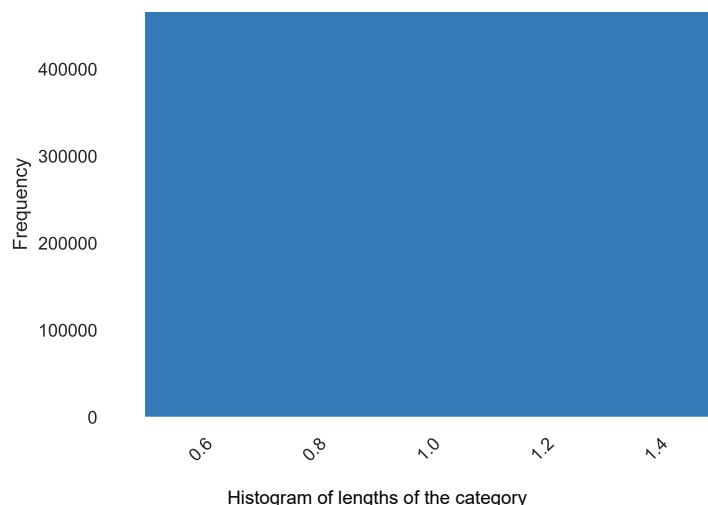
Sample

1st row	B
2nd row	C
3rd row	C
4th row	C
5th row	B

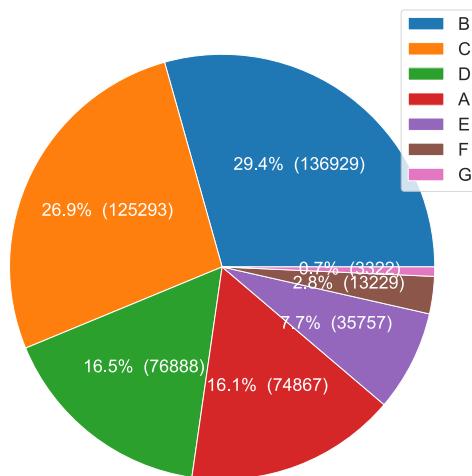
Common Values

Value	Count	Frequency (%)
B	136929	29.4%
C	125293	26.9%
D	76888	16.5%
A	74867	16.1%
E	35757	7.7%
F	13229	2.8%
G	3322	0.7%

Length



Pie chart



Value	Count	Frequency (%)
b	136929	29.4%
c	125293	26.9%
d	76888	16.5%
a	74867	16.1%
e	35757	7.7%
f	13229	2.8%
g	3322	0.7%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

sub_grade

Categorical

HIGH

CORRELATION (This variable has a high correlation with 3 fields: grade, policy_code, application_type)

HIGH CORRELATION (This variable has a high correlation with 3 fields: term, int_rate, grade)

Distinct	35
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	2
Median length	2
Mean length	2
Min length	2

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

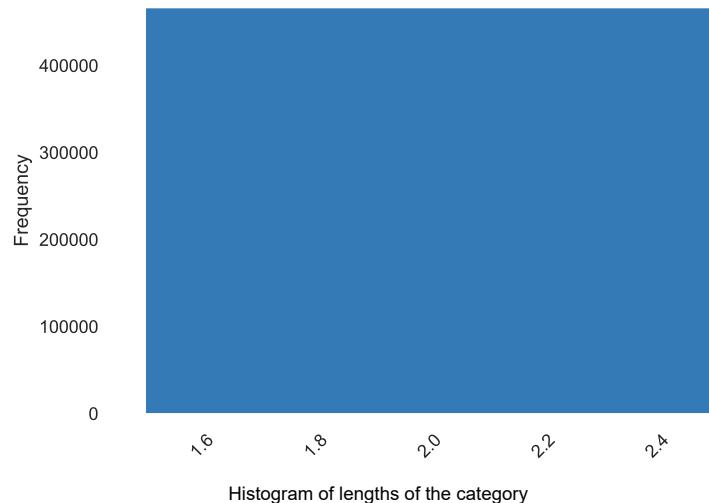
Sample

1st row	B2
2nd row	C4
3rd row	C5
4th row	C1
5th row	B5

Common Values

Value	Count	Frequency (%)
B3	31686	6.8%
B4	30505	6.5%
C1	26953	5.8%
C2	26740	5.7%
B2	26610	5.7%
C3	25317	5.4%
B5	25252	5.4%
C4	24105	5.2%
B1	22876	4.9%
C5	22178	4.8%
Other values (25)	204063	43.8%

Length



Value	Count	Frequency (%)
b3	31686	6.8%
b4	30505	6.5%
c1	26953	5.8%
c2	26740	5.7%
b2	26610	5.7%
c3	25317	5.4%
b5	25252	5.4%
c4	24105	5.2%
b1	22876	4.9%
c5	22178	4.8%
Other values (25)	204063	43.8%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

emp_title

Categorical

HIGH

CARDINALITY

MISSING

Distinct

205475

Distinct (%)

46.8%

Missing

27588

Missing (%)

5.9%

Memory size

3.6 MiB

Length

Max length	78
Median length	16
Mean length	16.75280889
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	172551
Unique (%)	39.3%

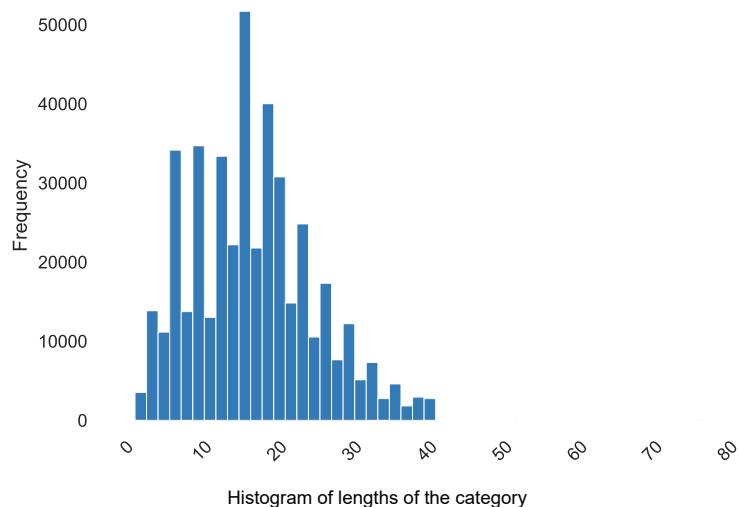
Sample

1st row	Ryder
2nd row	AIR RESOURCES BOARD
3rd row	University Medical Group
4th row	Veolia Transportaton
5th row	Southern Star Photography

Common Values

Value	Count	Frequency (%)
Teacher	5399	1.2%
Manager	4438	1.0%
Registered Nurse	2316	0.5%
RN	2204	0.5%
Supervisor	1967	0.4%
Sales	1624	0.3%
Project Manager	1624	0.3%
Owner	1527	0.3%
Office Manager	1395	0.3%
manager	1312	0.3%
Other values (205465)	414891	89.0%
(Missing)	27588	5.9%

Length



Value	Count	Frequency (%)
manager	42458	4.2%
of	20198	2.0%
inc	13828	1.4%
director	11210	1.1%
sales	10235	1.0%
assistant	10042	1.0%
	9076	0.9%
supervisor	8065	0.8%
teacher	8010	0.8%
service	7973	0.8%
Other values (64613)	865783	86.0%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

emp_length

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

MISSING

Distinct	11
Distinct (%)	< 0.1%
Missing	21008
Missing (%)	4.5%
Memory size	3.6 MiB

Length

Max length	9
Median length	7
Mean length	7.688876812
Min length	6

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

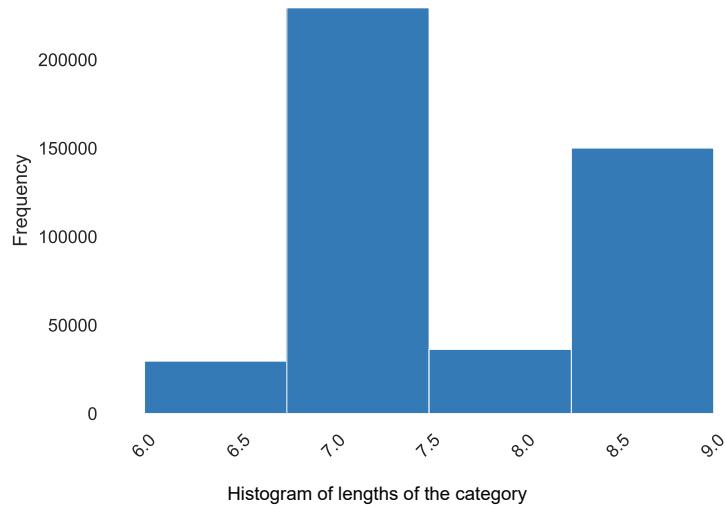
Sample

1st row	10+ years
2nd row	< 1 year
3rd row	10+ years
4th row	10+ years
5th row	1 year

Common Values

Value	Count	Frequency (%)
10+ years	150049	32.2%
2 years	41373	8.9%
3 years	36596	7.8%
< 1 year	36265	7.8%
5 years	30774	6.6%
1 year	29622	6.4%
4 years	28023	6.0%
7 years	26180	5.6%
6 years	26112	5.6%
8 years	22395	4.8%
(Missing)	21008	4.5%

Length



Value	Count	Frequency (%)
years	379390	40.9%
10	150049	16.2%
1	65887	7.1%
year	65887	7.1%
2	41373	4.5%
3	36596	3.9%
	36265	3.9%
5	30774	3.3%
4	28023	3.0%
7	26180	2.8%
Other values (3)	66395	7.2%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

home_ownership

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

Distinct	6
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	8
Median length	8
Mean length	5.934389912
Min length	3

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	1	?
Unique (%)	< 0.1%	

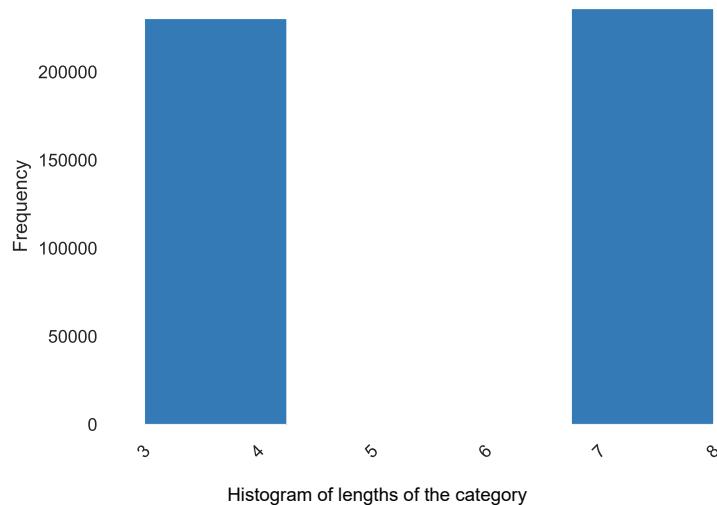
Sample

1st row	RENT
2nd row	RENT
3rd row	RENT
4th row	RENT
5th row	RENT

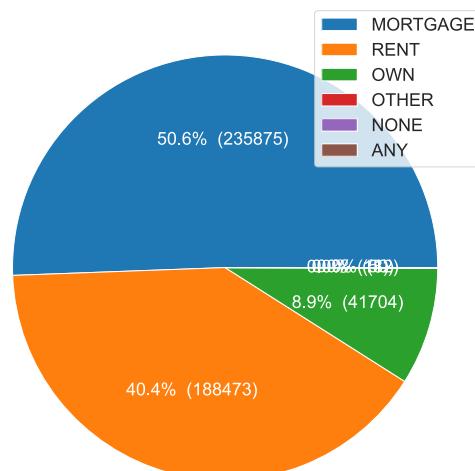
Common Values

Value	Count	Frequency (%)
MORTGAGE	235875	50.6%
RENT	188473	40.4%
OWN	41704	8.9%
OTHER	182	< 0.1%
NONE	50	< 0.1%
ANY	1	< 0.1%

Length



Pie chart



Value	Count	Frequency (%)
mortgage	235875	50.6%
rent	188473	40.4%
own	41704	8.9%
other	182	< 0.1%
none	50	< 0.1%
any	1	< 0.1%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

annual_inc

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: tot.cur bal)

SKEWED

Distinct	31901
Distinct (%)	6.8%
Missing	4
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	73277.38147
Minimum	1896
Maximum	7500000
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



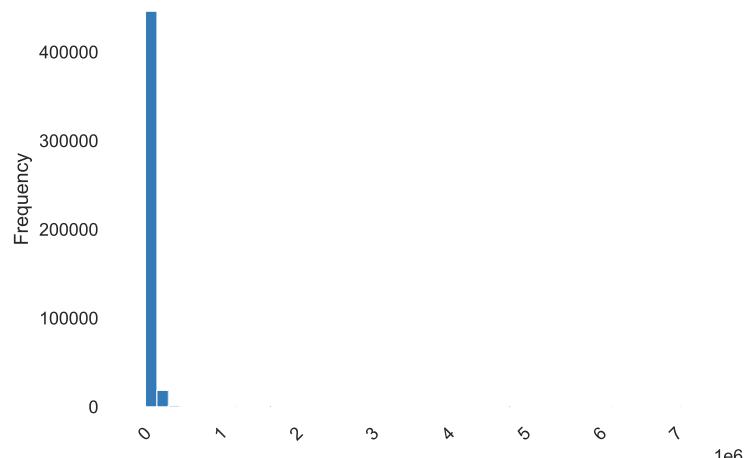
Quantile statistics

Minimum	1896
5-th percentile	28000
Q1	45000
median	63000
Q3	88960
95-th percentile	150000
Maximum	7500000
Range	7498104
Interquartile range (IQR)	43960

Descriptive statistics

Standard deviation	54963.56865
Coefficient of variation (CV)	0.7500755015
Kurtosis	3071.11437
Mean	73277.38147
Median Absolute Deviation (MAD)	21000
Skewness	29.56890538

Sum	$3.416785071 \times 10^{10}$
Variance	3020993879
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
60000	17980	3.9%
50000	15949	3.4%
65000	13211	2.8%
40000	12864	2.8%
70000	12436	2.7%
45000	11986	2.6%
80000	11641	2.5%
75000	11509	2.5%
55000	10938	2.3%
90000	8944	1.9%
Other values (31891)	338823	72.7%

Value	Count	Frequency (%)
1896	1	< 0.1%
2000	1	< 0.1%
3000	1	< 0.1%
3300	1	< 0.1%
3500	1	< 0.1%
3600	1	< 0.1%
4000	5	< 0.1%
4080	1	< 0.1%
4200	2	< 0.1%
4500	1	< 0.1%

Value	Count	Frequency (%)
7500000	1	< 0.1%
7446395	1	< 0.1%
7141778	1	< 0.1%
6100000	1	< 0.1%
6000000	1	< 0.1%
5000000	1	< 0.1%
4900000	1	< 0.1%
4560077	1	< 0.1%
3900000	1	< 0.1%
2039784	1	< 0.1%

verification_status

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	15
Median length	12
Mean length	11.52337948
Min length	8

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

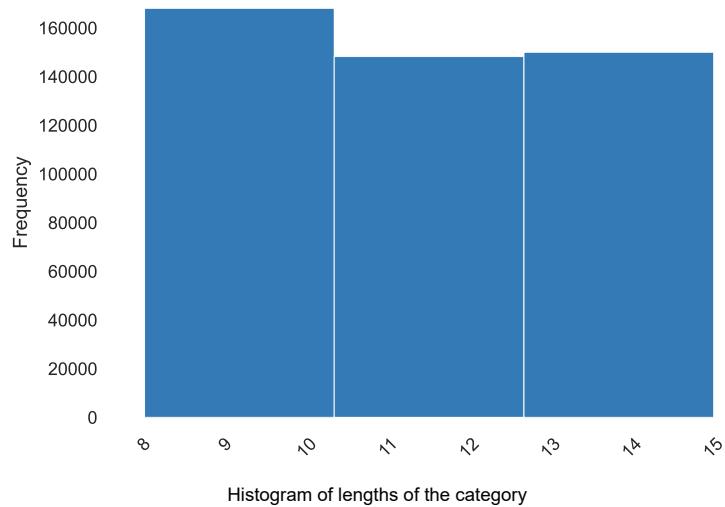
Sample

1st row	Verified
2nd row	Source Verified
3rd row	Not Verified
4th row	Source Verified
5th row	Source Verified

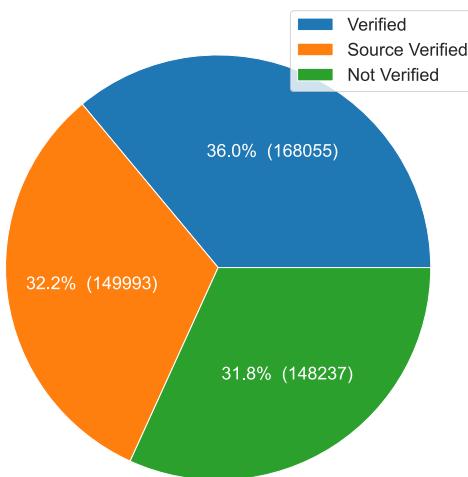
Common Values

Value	Count	Frequency (%)
Verified	168055	36.0%
Source Verified	149993	32.2%
Not Verified	148237	31.8%

Length



Pie chart



Value	Count	Frequency (%)
verified	466285	61.0%
source	149993	19.6%
not	148237	19.4%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

issue_d

Categorical

HIGH

CARDINALITY

HIGH CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

HIGH CORRELATION (This variable has a high correlation with 8 fields: Unnamed: 0, id, member_id, loan_status, mths_since_last_record, initial_list_status, last_pymnt_d, next_pymnt_d)

Distinct	91
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	6
Median length	6
Mean length	6
Min length	6

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

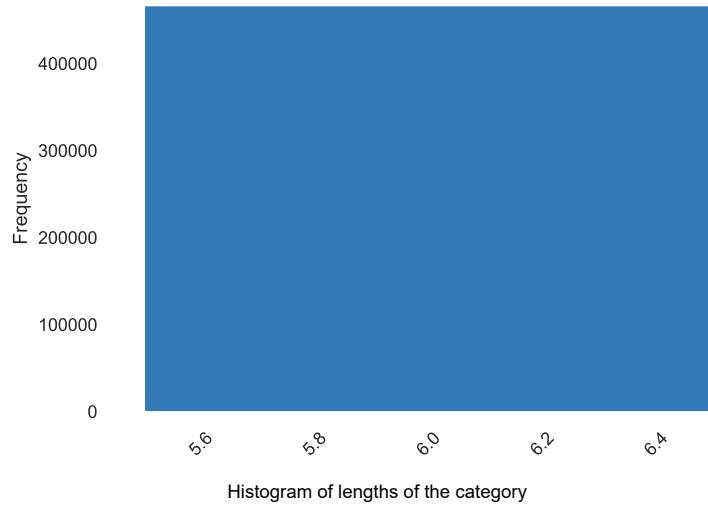
Sample

1st row	Dec-11
2nd row	Dec-11
3rd row	Dec-11
4th row	Dec-11
5th row	Dec-11

Common Values

Value	Count	Frequency (%)
Oct-14	38782	8.3%
Jul-14	29306	6.3%
Nov-14	25054	5.4%
May-14	19099	4.1%
Apr-14	19071	4.1%
Aug-14	18814	4.0%
Jun-14	17179	3.7%
Mar-14	16513	3.5%
Jan-14	15628	3.4%
Feb-14	15269	3.3%
Other values (81)	251570	54.0%

Length



Value	Count	Frequency (%)
oct-14	38782	8.3%
jul-14	29306	6.3%
nov-14	25054	5.4%
may-14	19099	4.1%
apr-14	19071	4.1%
aug-14	18814	4.0%
jun-14	17179	3.7%
mar-14	16513	3.5%
jan-14	15628	3.4%
feb-14	15269	3.3%
Other values (81)	251570	54.0%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

loan_status

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

HIGH CORRELATION (This variable has a high correlation with 5 fields: issue_d, out_prncp, out_prncp_inv, last_pymnt_d, next_pymnt_d)

Distinct	9
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	51
Median length	10
Mean length	9.050962394
Min length	7

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

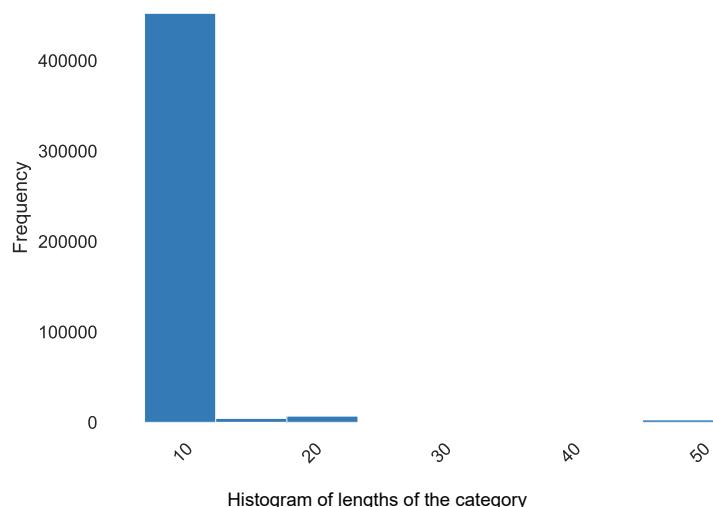
Sample

1st row	Fully Paid
2nd row	Charged Off
3rd row	Fully Paid
4th row	Fully Paid
5th row	Current

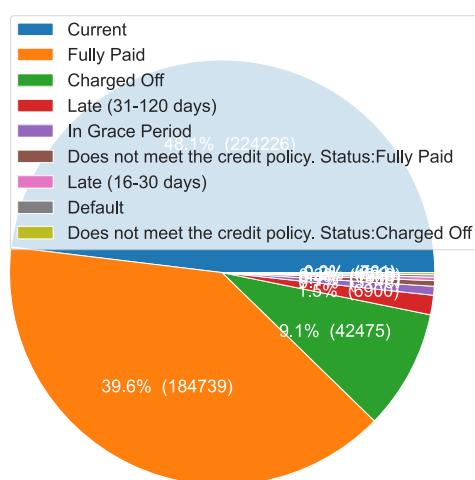
Common Values

Value	Count	Frequency (%)
Current	224226	48.1%
Fully Paid	184739	39.6%
Charged Off	42475	9.1%
Late (31-120 days)	6900	1.5%
In Grace Period	3146	0.7%
Does not meet the credit policy. Status:Fully Paid	1988	0.4%
Late (16-30 days)	1218	0.3%
Default	832	0.2%
Does not meet the credit policy. Status:Charged Off	761	0.2%

Length



Pie chart



Value	Count	Frequency (%)
current	224226	30.5%
paid	186727	25.4%
fully	184739	25.1%
off	43236	5.9%
charged	42475	5.8%
late	8118	1.1%
days	8118	1.1%
31-120	6900	0.9%
grace	3146	0.4%
period	3146	0.4%
Other values (11)	24439	3.3%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

pymnt_plan

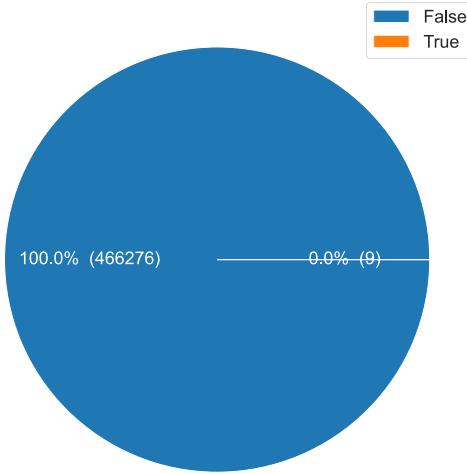
Boolean

HIGH

CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	455.5 KiB

Value	Count	Frequency (%)
False	466276	> 99.9%
True	9	< 0.1%



url	
Categorical	
CARDINALITY	HIGH
UNIFORM	
UNIQUE	
Distinct	466285
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	69
Median length	69
Mean length	68.42436922
Min length	66

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	466285	?
Unique (%)	100.0%	

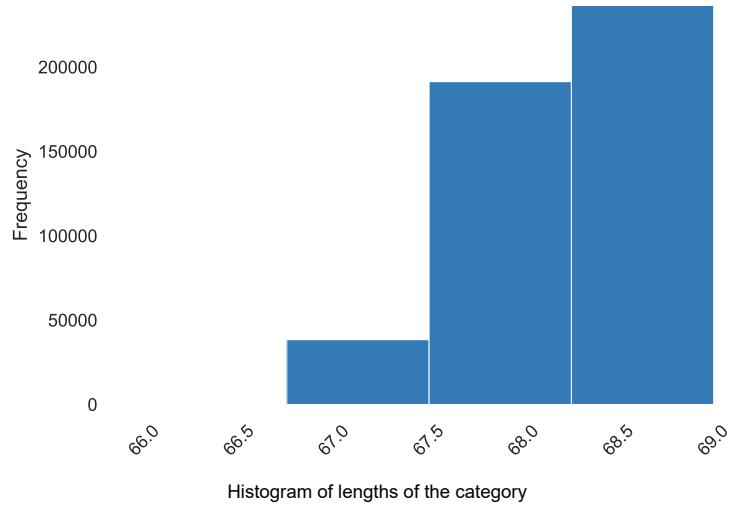
Sample

1st row	https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077501
2nd row	https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077430
3rd row	https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077175
4th row	https://www.lendingclub.com/browse/loanDetail.action?loan_id=1076863
5th row	https://www.lendingclub.com/browse/loanDetail.action?loan_id=1075358

Common Values

Value	Count	Frequency (%)
https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077501	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26180404	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26219322	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26239577	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26290537	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26230483	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26350584	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26269915	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26340359	1	< 0.1%
https://www.lendingclub.com/browse/loanDetail.action?loan_id=26299926	1	< 0.1%
Other values (466275)	466275	> 99.9%

Length



Value	Count	Frequency (%)
https://www.lendingclub.com/browse/loandetail.action?loan_id=1077501	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1068409	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1076863	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1075358	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1075269	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1069639	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1072053	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1071795	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1071570	1	< 0.1%
https://www.lendingclub.com/browse/loandetail.action?loan_id=1070078	1	< 0.1%
Other values (466275)	466275	> 99.9%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

desc	
Categorical	HIGH
CARDINALITY	
MISSING	
UNIFORM	
Distinct	124436
Distinct (%)	98.8%
Missing	340302
Missing (%)	73.0%
Memory size	3.6 MiB

Length

Max length	3988
Median length	165
Mean length	239.9754491
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	123600
Unique (%)	98.1%

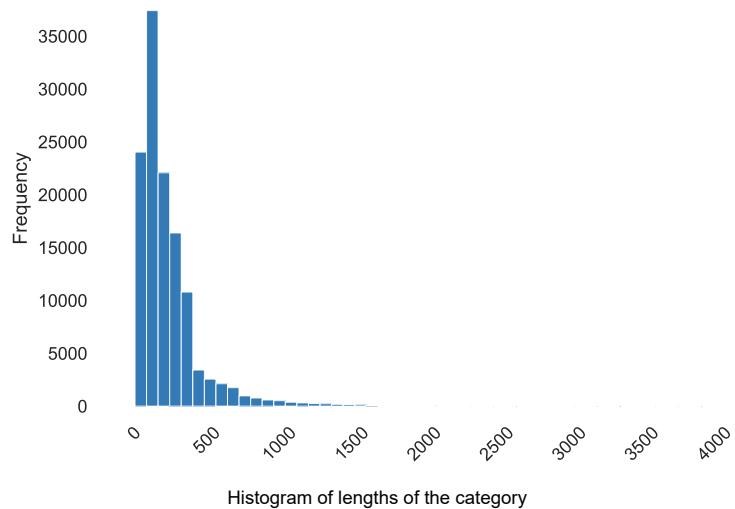
Sample

1st Borrower added on 12/22/11 > I need to upgrade my business technologies. row
2nd Borrower added on 12/22/11 > I plan to use this money to finance the motorcycle i am looking at. I plan to have it paid off as soon as possible/when i sell my old row bike. I only need this money because the deal im looking at is to good to pass up. Borrower added on 12/22/11 > I plan to use this money to finance the motorcycle i am looking at. I plan to have it paid off as soon as possible/when i sell my old bike.I only need this money because the deal im looking at is to good to pass up. I have finished college with an associates degree in business and its taking me places
3rd Borrower added on 12/21/11 > to pay for property tax (borrow from friend, need to pay back) & central A/C need to be replace. I'm very sorry to let my loan expired row last time.
4th Borrower added on 12/21/11 > I plan on combining three large interest bills together and freeing up some extra each month to pay toward other bills. I've always row been a good payor but have found myself needing to make adjustments to my budget due to a medical scare. My job is very stable, I love it.
5th Borrower added on 12/18/11 > I am planning on using the funds to pay off two retail credit cards with 24.99% interest rates, as well as a major bank credit card row with a 18.99% rate. I pay all my bills on time, looking for a lower combined payment and lower monthly payment.

Common Values

Value	Count	Frequency (%)
	234	0.1%
Debt Consolidation	13	< 0.1%
Borrower added on 03/17/14 > Debt consolidation 	11	< 0.1%
Borrower added on 03/10/14 > Debt consolidation 	10	< 0.1%
Borrower added on 02/19/14 > Debt consolidation 	9	< 0.1%
Borrower added on 01/29/14 > Debt consolidation 	8	< 0.1%
Camping Membership	8	< 0.1%
Borrower added on 01/22/14 > Debt consolidation 	7	< 0.1%
Borrower added on 01/15/14 > Debt consolidation 	7	< 0.1%
Borrower added on 03/05/14 > debt consolidation 	6	< 0.1%
Other values (124426)	125670	27.0%
(Missing)	340302	73.0%

Length



Value	Count	Frequency (%)
to	207625	3.8%
on	187307	3.5%
i	181919	3.4%
	156800	2.9%
added	146492	2.7%
borrower	146417	2.7%
and	138078	2.6%
my	132657	2.5%
a	124382	2.3%
the	114384	2.1%
Other values (85640)	3856875	71.5%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

purpose

Categorical

HIGH**CORRELATION** (This variable has a high correlation with 2 fields: policy_code, application_type)

Distinct	14
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	18
Median length	18
Mean length	14.97865469
Min length	3

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

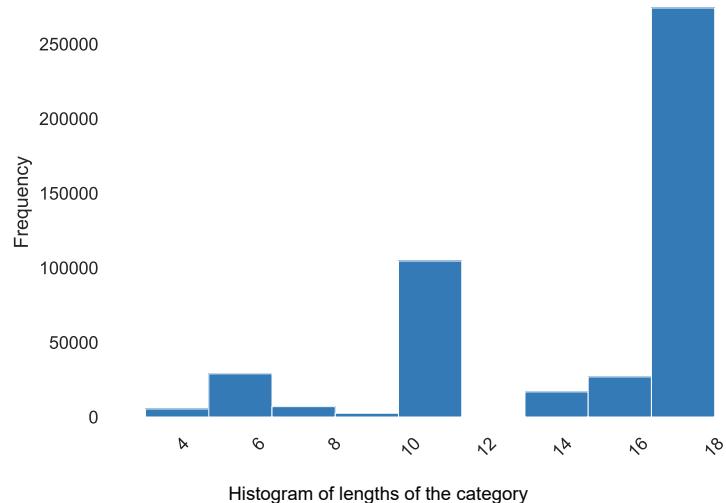
Sample

1st row	credit_card
2nd row	car
3rd row	small_business
4th row	other
5th row	other

Common Values

Value	Count	Frequency (%)
debt_consolidation	274195	58.8%
credit_card	104157	22.3%
home_improvement	26537	5.7%
other	23690	5.1%
major_purchase	9828	2.1%
small_business	7013	1.5%
car	5397	1.2%
medical	4602	1.0%
moving	2994	0.6%
vacation	2487	0.5%
Other values (4)	5385	1.2%

Length



Value	Count	Frequency (%)
debt_consolidation	274195	58.8%
credit_card	104157	22.3%
home_improvement	26537	5.7%
other	23690	5.1%
major_purchase	9828	2.1%
small_business	7013	1.5%
car	5397	1.2%
medical	4602	1.0%
moving	2994	0.6%
vacation	2487	0.5%
Other values (4)	5385	1.2%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

title
Categorical

HIGH

CARDINALITY

Distinct	63099
Distinct (%)	13.5%
Missing	20
Missing (%)	< 0.1%
Memory size	3.6 MiB

Length

Max length	80
Median length	18
Mean length	17.28925397
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	53923
Unique (%)	11.6%

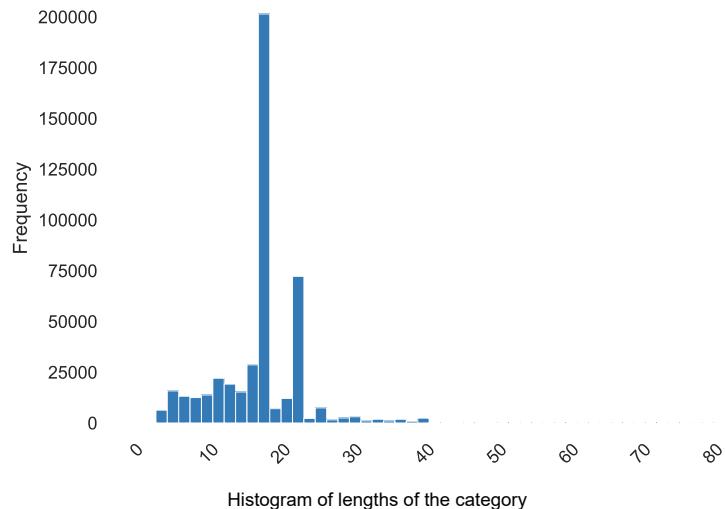
Sample

1st row	Computer
2nd row	bike
3rd row	real estate business
4th row	personel
5th row	Personal

Common Values

Value	Count	Frequency (%)
Debt consolidation	164075	35.2%
Credit card refinancing	62275	13.4%
Debt Consolidation	15760	3.4%
Home improvement	14849	3.2%
Other	12713	2.7%
Consolidation	5381	1.2%
debt consolidation	4834	1.0%
Major purchase	4617	1.0%
Debt Consolidation Loan	3804	0.8%
Business	3371	0.7%
Other values (63089)	174586	37.4%

Length



Value	Count	Frequency (%)
consolidation	216552	20.5%
debt	216162	20.4%
credit	92756	8.8%
card	84588	8.0%
refinancing	63296	6.0%
loan	37599	3.6%
home	24162	2.3%
improvement	19653	1.9%
other	13157	1.2%
payoff	8966	0.8%
Other values (17793)	280651	26.5%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

zip_code

Categorical

HIGH

CARDINALITY

Distinct

888

Distinct (%)

0.2%

Missing

0

Missing (%)

0.0%

Memory size

3.6 MiB

Length

Max length	5
Median length	5
Mean length	5
Min length	5

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	30	
Unique (%)	< 0.1%	

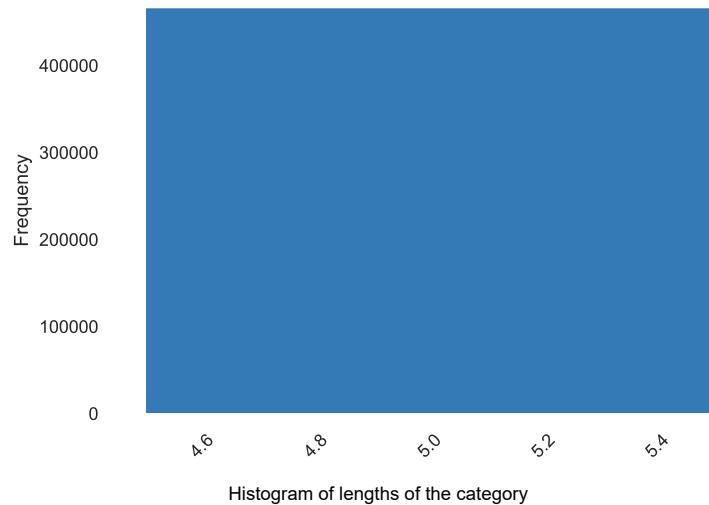
Sample

1st row	860xx	
2nd row	309xx	
3rd row	606xx	
4th row	917xx	
5th row	972xx	

Common Values

Value	Count	Frequency (%)
945xx	5304	1.1%
112xx	5102	1.1%
750xx	5013	1.1%
606xx	4696	1.0%
100xx	4391	0.9%
300xx	4120	0.9%
900xx	4069	0.9%
070xx	4067	0.9%
331xx	3983	0.9%
770xx	3587	0.8%
Other values (878)	421953	90.5%

Length



Value	Count	Frequency (%)
945xx	5304	1.1%
112xx	5102	1.1%
750xx	5013	1.1%
606xx	4696	1.0%
100xx	4391	0.9%
300xx	4120	0.9%
900xx	4069	0.9%
070xx	4067	0.9%
331xx	3983	0.9%
770xx	3587	0.8%
Other values (878)	421953	90.5%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

addr_state

Categorical

HIGH**CORRELATION** (This variable has a high correlation with 2 fields: policy_code, application_type)

Distinct	50
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	2
Median length	2
Mean length	2
Min length	2

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

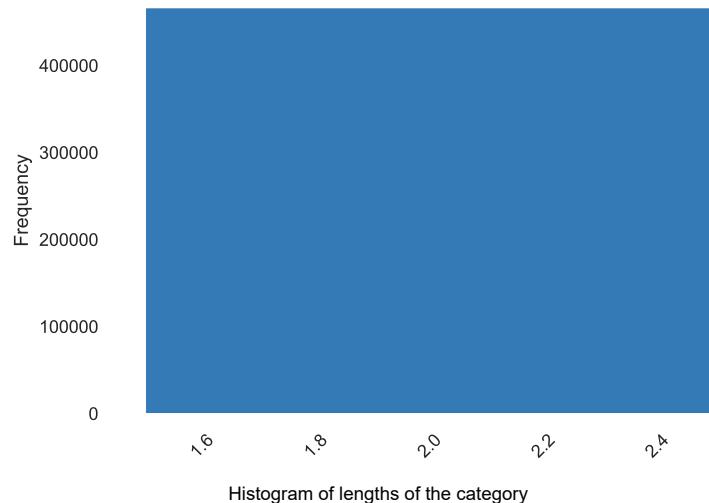
Sample

1st row	AZ
2nd row	GA
3rd row	IL
4th row	CA
5th row	OR

Common Values

Value	Count	Frequency (%)
CA	71450	15.3%
NY	40242	8.6%
TX	36439	7.8%
FL	31637	6.8%
IL	18612	4.0%
NJ	18061	3.9%
PA	16424	3.5%
OH	15237	3.3%
GA	14975	3.2%
VA	14222	3.1%
Other values (40)	188986	40.5%

Length



Value	Count	Frequency (%)
ca	71450	15.3%
ny	40242	8.6%
tx	36439	7.8%
fl	31637	6.8%
il	18612	4.0%
nj	18061	3.9%
pa	16424	3.5%
oh	15237	3.3%
ga	14975	3.2%
va	14222	3.1%
Other values (40)	188986	40.5%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

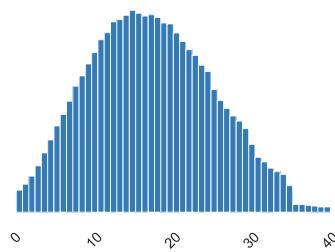
Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

dtiReal number ($\mathbb{R}_{\geq 0}$)

Distinct	3997
Distinct (%)	0.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	17.21875793
Minimum	0
Maximum	39.99
Zeros	363
Zeros (%)	0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB

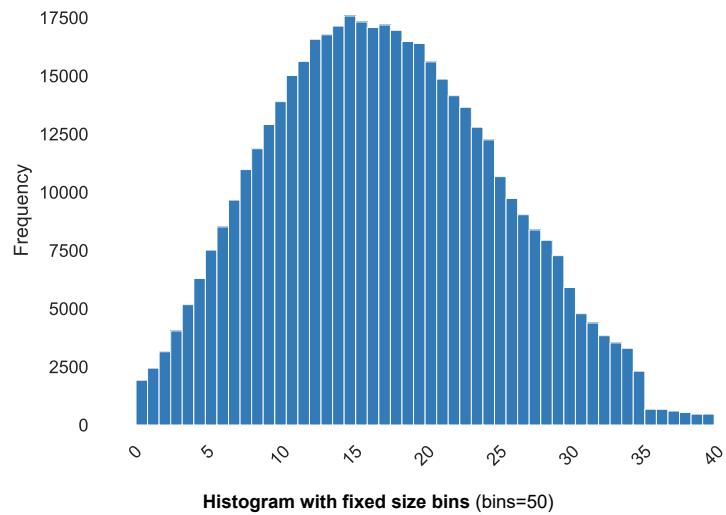
**Quantile statistics**

Minimum	0
5-th percentile	4.82
Q1	11.36
median	16.87
Q3	22.78
95-th percentile	30.77
Maximum	39.99
Range	39.99
Interquartile range (IQR)	11.42

Descriptive statistics

Standard deviation	7.851121209
Coefficient of variation (CV)	0.4559632723
Kurtosis	-0.5407010003
Mean	17.21875793
Median Absolute Deviation (MAD)	5.7
Skewness	0.1882832551
Sum	8028848.54

Variance	61.64010424
Monotonicity	Not monotonic



Value	Count	Frequency (%)
14.4	410	0.1%
19.2	381	0.1%
12	369	0.1%
16.8	368	0.1%
18	365	0.1%
13.2	364	0.1%
0	363	0.1%
20.4	354	0.1%
15.6	341	0.1%
21.6	308	0.1%
Other values (3987)	462662	99.2%

Value	Count	Frequency (%)
0	363	0.1%
0.01	9	< 0.1%
0.02	9	< 0.1%
0.03	4	< 0.1%
0.04	5	< 0.1%
0.05	4	< 0.1%
0.06	5	< 0.1%
0.07	9	< 0.1%
0.08	9	< 0.1%
0.09	6	< 0.1%

Value	Count	Frequency (%)
39.99	8	< 0.1%
39.98	5	< 0.1%
39.97	6	< 0.1%
39.96	4	< 0.1%
39.95	5	< 0.1%
39.94	1	< 0.1%
39.93	5	< 0.1%
39.92	9	< 0.1%
39.91	7	< 0.1%
39.9	5	< 0.1%

delinq_2yrs

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

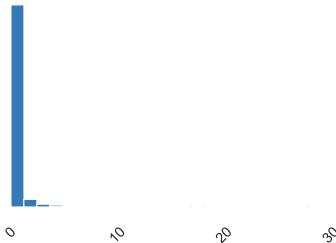
CORRELATION (This variable has a high correlation with 2 fields: mths since last delinq, mths since last major derog.)

HIGH CORRELATION (This variable has a high correlation with 1 fields: mths since last delinq)

HIGH CORRELATION (This variable has a high correlation with 1 fields: mths since last delinq)

ZEROS

Distinct	24
Distinct (%)	< 0.1%
Missing	29
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	0.2846783741
Minimum	0
Maximum	29
Zeros	382954
Zeros (%)	82.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



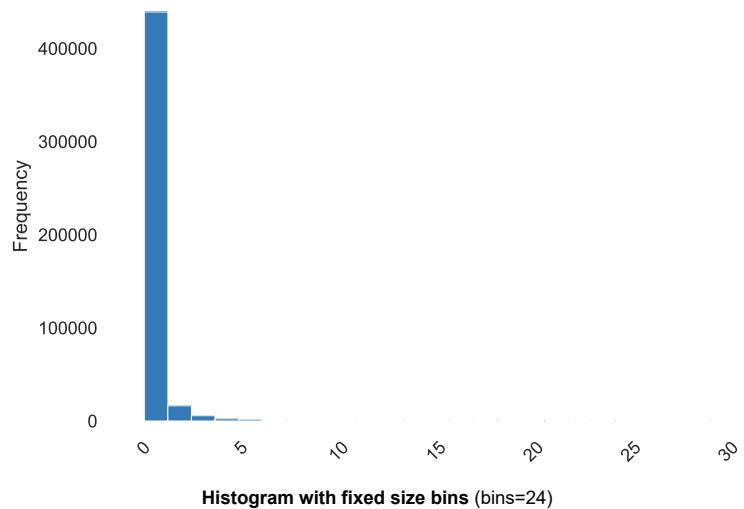
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	2
Maximum	29
Range	29
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.7973651064
Coefficient of variation (CV)	2.80093319
Kurtosis	53.14391843
Mean	0.2846783741

Median Absolute Deviation (MAD)	0
Skewness	5.34574792
Sum	132733
Variance	0.635791113
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	382954	82.1%
1	56224	12.1%
2	16310	3.5%
3	5684	1.2%
4	2378	0.5%
5	1236	0.3%
6	629	0.1%
7	346	0.1%
8	177	< 0.1%
9	111	< 0.1%
Other values (14)	207	< 0.1%

Value	Count	Frequency (%)
0	382954	82.1%
1	56224	12.1%
2	16310	3.5%
3	5684	1.2%
4	2378	0.5%
5	1236	0.3%
6	629	0.1%
7	346	0.1%
8	177	< 0.1%
9	111	< 0.1%

Value	Count	Frequency (%)
29	1	< 0.1%
24	1	< 0.1%
22	2	< 0.1%
21	2	< 0.1%
19	3	< 0.1%
18	6	< 0.1%
17	3	< 0.1%
16	5	< 0.1%
15	7	< 0.1%
14	8	< 0.1%

earliest_cr_line

Categorical

HIGH

CARDINALITY

Distinct

664

Distinct (%)

0.1%

Missing

29

Missing (%)

< 0.1%

Memory size

3.6 MiB

Length

Max length	6
Median length	6
Mean length	6
Min length	6

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	34	
Unique (%)	< 0.1%	

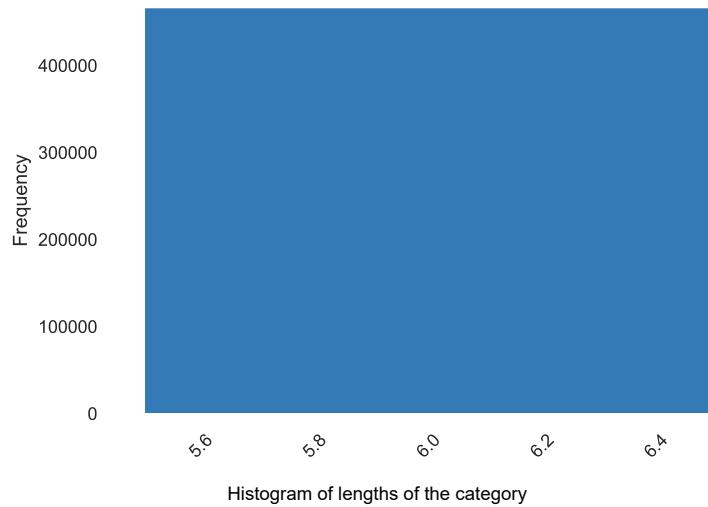
Sample

1st row	Jan-85	
2nd row	Apr-99	
3rd row	Nov-01	
4th row	Feb-96	
5th row	Jan-96	

Common Values

Value	Count	Frequency (%)
Oct-00	3674	0.8%
Aug-00	3597	0.8%
Aug-01	3455	0.7%
Oct-99	3406	0.7%
Oct-01	3329	0.7%
Nov-99	3323	0.7%
Nov-00	3267	0.7%
Sep-00	3256	0.7%
Sep-99	3130	0.7%
Aug-99	3126	0.7%
Other values (654)	432693	92.8%

Length



Value	Count	Frequency (%)
oct-00	3674	0.8%
aug-00	3597	0.8%
aug-01	3455	0.7%
oct-99	3406	0.7%
oct-01	3329	0.7%
nov-99	3323	0.7%
nov-00	3267	0.7%
sep-00	3256	0.7%
sep-99	3130	0.7%
aug-99	3126	0.7%
Other values (654)	432693	92.8%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

inq_last_6mths

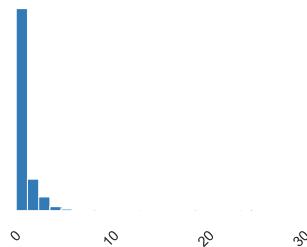
Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: next_pymnt_d)

ZEROS

Distinct	28
Distinct (%)	< 0.1%
Missing	29
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	0.8047446038
Minimum	0
Maximum	33
Zeros	241498
Zeros (%)	51.8%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



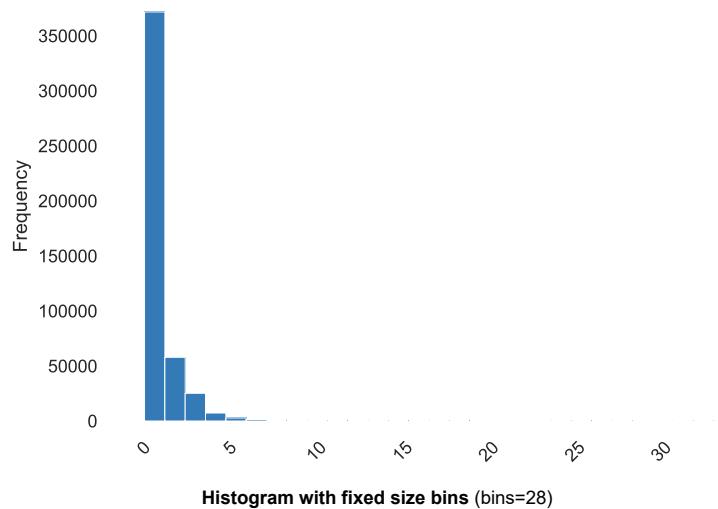
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	1
95-th percentile	3
Maximum	33
Range	33
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	1.091598121
Coefficient of variation (CV)	1.356452862
Kurtosis	13.86388292
Mean	0.8047446038
Median Absolute Deviation (MAD)	0
Skewness	2.142143376

Sum	375217
Variance	1.191586458
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	241498	51.8%
1	130148	27.9%
2	57825	12.4%
3	25126	5.4%
4	7241	1.6%
5	2768	0.6%
6	1188	0.3%
7	195	< 0.1%
8	122	< 0.1%
9	50	< 0.1%
Other values (18)	95	< 0.1%
(Missing)	29	< 0.1%

Value	Count	Frequency (%)
0	241498	51.8%
1	130148	27.9%
2	57825	12.4%
3	25126	5.4%
4	7241	1.6%
5	2768	0.6%
6	1188	0.3%
7	195	< 0.1%
8	122	< 0.1%
9	50	< 0.1%

Value	Count	Frequency (%)
33	1	< 0.1%
32	1	< 0.1%
31	1	< 0.1%
28	1	< 0.1%
27	1	< 0.1%
25	1	< 0.1%
24	2	< 0.1%
20	1	< 0.1%
19	2	< 0.1%
18	4	< 0.1%

mths_since_last_delinq

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: delinq_2yrs, mths_since_last_major_derog)

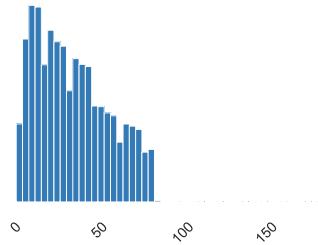
HIGH CORRELATION (This variable has a high correlation with 2 fields: delinq_2yrs, mths_since_last_major_derog)

HIGH CORRELATION (This variable has a high correlation with 2 fields: delinq_2yrs, mths_since_last_major_derog)

HIGH CORRELATION (This variable has a high correlation with 1 fields: mths_since_last_major_derog)

MISSING

Distinct	145
Distinct (%)	0.1%
Missing	250351
Missing (%)	53.7%
Infinite	0
Infinite (%)	0.0%
Mean	34.10443006
Minimum	0
Maximum	188
Zeros	1223
Zeros (%)	0.3%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



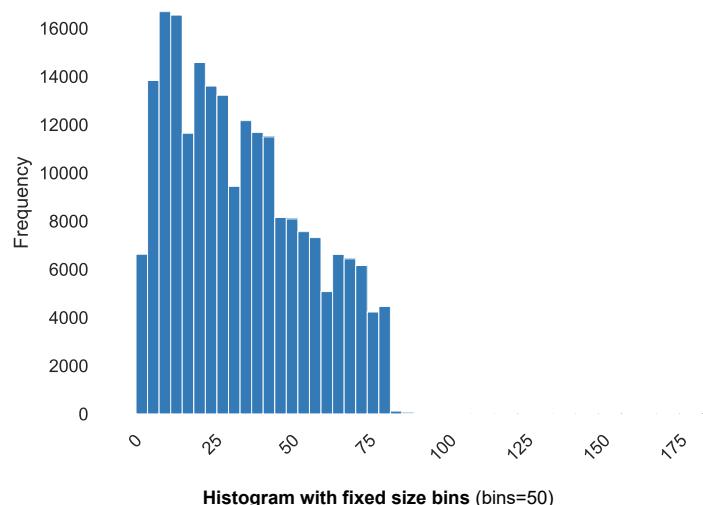
Quantile statistics

Minimum	0
5-th percentile	5
Q1	16
median	31
Q3	49
95-th percentile	74
Maximum	188
Range	188
Interquartile range (IQR)	33

Descriptive statistics

Standard deviation	21.77848702
Coefficient of variation (CV)	0.638582348
Kurtosis	-0.7323044779
Mean	34.10443006

Median Absolute Deviation (MAD)	17
Skewness	0.4645231113
Sum	7364306
Variance	474.3024971
Monotonicity	Not monotonic



Value	Count	Frequency (%)
9	4358	0.9%
8	4272	0.9%
13	4260	0.9%
12	4193	0.9%
10	4146	0.9%
14	4102	0.9%
7	4061	0.9%
6	4028	0.9%
15	3992	0.9%
16	3915	0.8%
Other values (135)	174607	37.4%
(Missing)	250351	53.7%

Value	Count	Frequency (%)
0	1223	0.3%
1	1414	0.3%
2	1824	0.4%
3	2164	0.5%
4	2611	0.6%
5	3134	0.7%
6	4028	0.9%
7	4061	0.9%
8	4272	0.9%
9	4358	0.9%

Value	Count	Frequency (%)
188	1	< 0.1%
180	1	< 0.1%
170	1	< 0.1%
159	1	< 0.1%
156	1	< 0.1%
152	1	< 0.1%
151	1	< 0.1%
149	1	< 0.1%
148	2	< 0.1%
146	1	< 0.1%

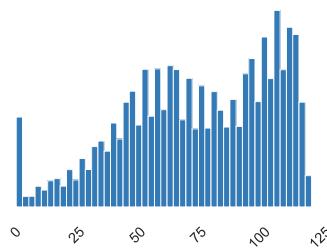
mths_since_last_record

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 4 fields: Unnamed: 0, issue d, last_pymnt_d, next_pymnt_d)
MISSING

Distinct	123
Distinct (%)	0.2%
Missing	403647
Missing (%)	86.6%
Infinite	0
Infinite (%)	0.0%
Mean	74.30601232
Minimum	0
Maximum	129
Zeros	1278
Zeros (%)	0.3%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



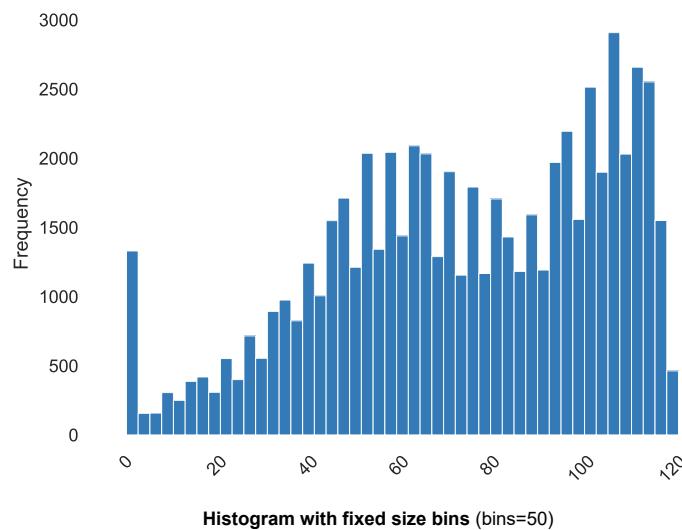
Quantile statistics

Minimum	0
5-th percentile	19
Q1	53
median	76
Q3	102
95-th percentile	115
Maximum	129
Range	129
Interquartile range (IQR)	49

Descriptive statistics

Standard deviation	30.35765285
Coefficient of variation (CV)	0.4085490784
Kurtosis	-0.6452106348
Mean	74.30601232
Median Absolute Deviation (MAD)	25
Skewness	-0.4484401122

Sum	4654380
Variance	921.5870868
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	1278	0.3%
109	1098	0.2%
108	999	0.2%
106	968	0.2%
104	953	0.2%
105	946	0.2%
107	941	0.2%
110	932	0.2%
112	910	0.2%
113	894	0.2%
Other values (113)	52719	11.3%
(Missing)	403647	86.6%

Value	Count	Frequency (%)
0	1278	0.3%
1	28	< 0.1%
2	24	< 0.1%
3	43	< 0.1%
4	51	< 0.1%
5	62	< 0.1%
6	68	< 0.1%
7	90	< 0.1%
8	93	< 0.1%
9	110	< 0.1%

Value	Count	Frequency (%)
129	1	< 0.1%
121	2	< 0.1%
120	2	< 0.1%
119	459	0.1%
118	777	0.2%
117	773	0.2%
116	821	0.2%
115	850	0.2%
114	881	0.2%
113	894	0.2%

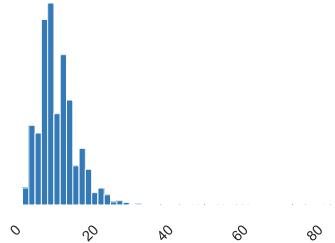
open_acc

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: total_acc)
HIGH CORRELATION (This variable has a high correlation with 1 fields: total_acc)
HIGH CORRELATION (This variable has a high correlation with 1 fields: total_acc)
HIGH CORRELATION (This variable has a high correlation with 1 fields: total_acc)

Distinct	62
Distinct (%)	< 0.1%
Missing	29
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	11.18706891
Minimum	0
Maximum	84
Zeros	7
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



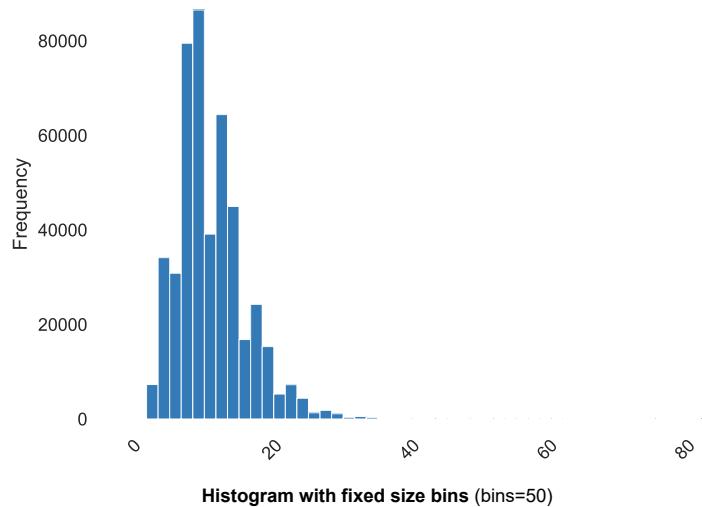
Quantile statistics

Minimum	0
5-th percentile	5
Q1	8
median	10
Q3	14
95-th percentile	20
Maximum	84
Range	84
Interquartile range (IQR)	6

Descriptive statistics

Standard deviation	4.987526124
Coefficient of variation (CV)	0.4458295703
Kurtosis	2.817051541
Mean	11.18706891

Median Absolute Deviation (MAD)	3
Skewness	1.165698165
Sum	5216038
Variance	24.87541683
Monotonicity	Not monotonic



Value	Count	Frequency (%)
9	44019	9.4%
10	42512	9.1%
8	41893	9.0%
11	39112	8.4%
7	37602	8.1%
12	34585	7.4%
6	30837	6.6%
13	29806	6.4%
14	24730	5.3%
5	21544	4.6%
Other values (52)	119616	25.7%

Value	Count	Frequency (%)
0	7	< 0.1%
1	142	< 0.1%
2	1790	0.4%
3	5500	1.2%
4	12588	2.7%
5	21544	4.6%
6	30837	6.6%
7	37602	8.1%
8	41893	9.0%
9	44019	9.4%

Value	Count	Frequency (%)
84	1	< 0.1%
76	2	< 0.1%
75	1	< 0.1%
62	1	< 0.1%
61	1	< 0.1%
58	4	< 0.1%
55	2	< 0.1%
54	2	< 0.1%
53	5	< 0.1%
52	2	< 0.1%

pub_rec

Real number ($\mathbb{R}_{\geq 0}$)

ZEROS

Distinct	26
Distinct (%)	< 0.1%
Missing	29
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	0.1605641536
Minimum	0
Maximum	63
Zeros	404893
Zeros (%)	86.8%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



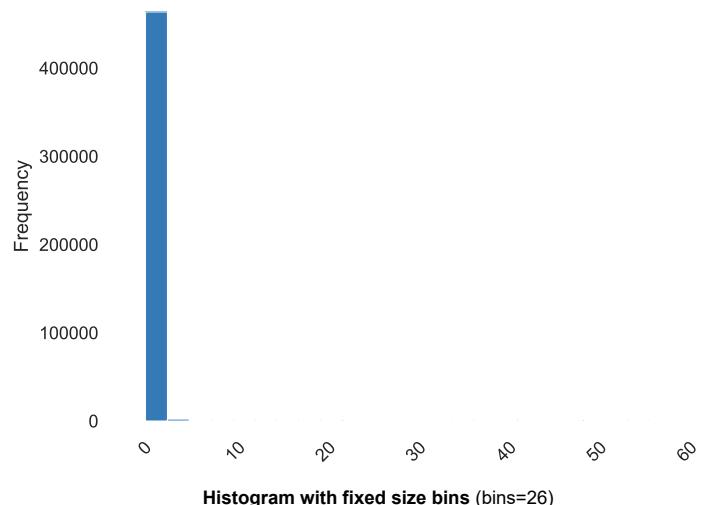
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	1
Maximum	63
Range	63
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.5108626365
Coefficient of variation (CV)	3.181673026
Kurtosis	1116.390912
Mean	0.1605641536
Median Absolute Deviation (MAD)	0
Skewness	14.9413546
Sum	74864

Variance	0.2609806334
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	404893	86.8%
1	53051	11.4%
2	5624	1.2%
3	1611	0.3%
4	520	0.1%
5	276	0.1%
6	136	< 0.1%
7	62	< 0.1%
8	29	< 0.1%
9	16	< 0.1%
Other values (16)	38	< 0.1%
(Missing)	29	< 0.1%

Value	Count	Frequency (%)
0	404893	86.8%
1	53051	11.4%
2	5624	1.2%
3	1611	0.3%
4	520	0.1%
5	276	0.1%
6	136	< 0.1%
7	62	< 0.1%
8	29	< 0.1%
9	16	< 0.1%

Value	Count	Frequency (%)
63	1	< 0.1%
54	1	< 0.1%
49	1	< 0.1%
40	1	< 0.1%
34	1	< 0.1%
21	1	< 0.1%
19	1	< 0.1%
18	2	< 0.1%
17	1	< 0.1%
16	1	< 0.1%

revol_bal

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

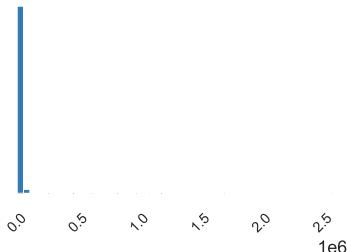
CORRELATION (This variable has a high correlation with 3 fields: loan_amnt, funded_amnt, total_rev_hi_lim)

HIGH CORRELATION (This variable has a high correlation with 1 fields: total_rev_hi_lim)

HIGH CORRELATION (This variable has a high correlation with 1 fields: total_rev_hi_lim)

HIGH CORRELATION (This variable has a high correlation with 2 fields: tot_cur_bal, total_rev_hi_lim)

Distinct	58142
Distinct (%)	12.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	16230.20349
Minimum	0
Maximum	2568995
Zeros	2269
Zeros (%)	0.5%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



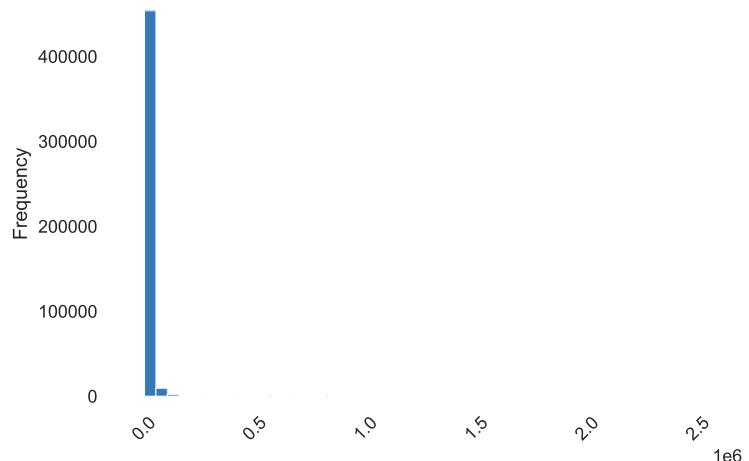
Quantile statistics

Minimum	0
5-th percentile	1955
Q1	6413
median	11764
Q3	20333
95-th percentile	40952.8
Maximum	2568995
Range	2568995
Interquartile range (IQR)	13920

Descriptive statistics

Standard deviation	20676.24515
Coefficient of variation (CV)	1.273936286
Kurtosis	1433.063059
Mean	16230.20349

Median Absolute Deviation (MAD)	6313
Skewness	19.4762908
Sum	7567900433
Variance	427507113.6
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	2269	0.5%
6052	40	< 0.1%
8696	39	< 0.1%
5896	39	< 0.1%
4532	39	< 0.1%
6223	39	< 0.1%
5365	39	< 0.1%
4427	39	< 0.1%
5655	39	< 0.1%
7429	39	< 0.1%
Other values (58132)	463664	99.4%

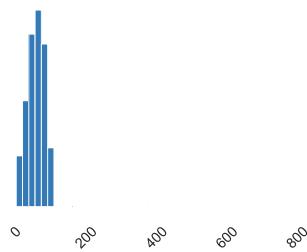
Value	Count	Frequency (%)
0	2269	0.5%
1	29	< 0.1%
2	25	< 0.1%
3	23	< 0.1%
4	19	< 0.1%
5	28	< 0.1%
6	22	< 0.1%
7	15	< 0.1%
8	13	< 0.1%
9	23	< 0.1%

Value	Count	Frequency (%)
2568995	1	< 0.1%
2560703	1	< 0.1%
1746716	1	< 0.1%
1743266	1	< 0.1%
1298783	1	< 0.1%
1207359	1	< 0.1%
1190046	1	< 0.1%
1030826	1	< 0.1%
988767	1	< 0.1%
975800	1	< 0.1%

revol_util

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	1269
Distinct (%)	0.3%
Missing	340
Missing (%)	0.1%
Infinite	0
Infinite (%)	0.0%
Mean	56.17694743
Minimum	0
Maximum	892.3
Zeros	2284
Zeros (%)	0.5%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



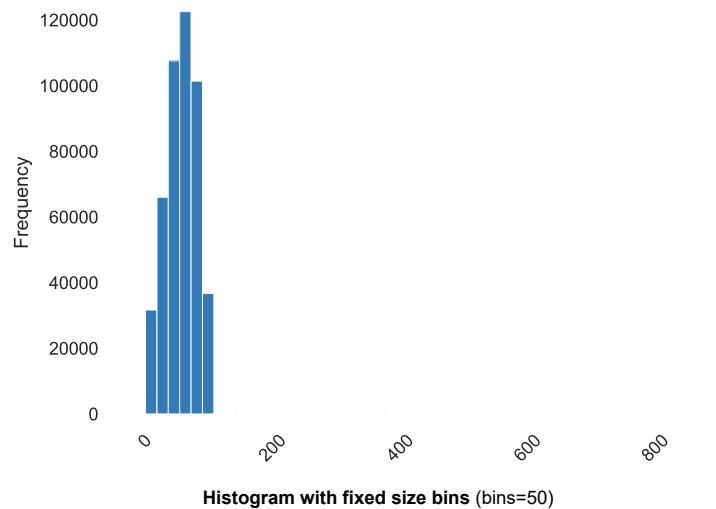
Quantile statistics

Minimum	0
5-th percentile	14.1
Q1	39.2
median	57.6
Q3	74.7
95-th percentile	92.5
Maximum	892.3
Range	892.3
Interquartile range (IQR)	35.5

Descriptive statistics

Standard deviation	23.73262798
Coefficient of variation (CV)	0.422462043
Kurtosis	2.691899262
Mean	56.17694743
Median Absolute Deviation (MAD)	17.7
Skewness	-0.160961515
Sum	26175367.77

Variance	563.2376308
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	2284	0.5%
58	864	0.2%
53	850	0.2%
59	845	0.2%
61	832	0.2%
57	822	0.2%
60	818	0.2%
63	817	0.2%
62	811	0.2%
65	809	0.2%
Other values (1259)	456193	97.8%

Value	Count	Frequency (%)
0	2284	0.5%
0.01	1	< 0.1%
0.03	1	< 0.1%
0.04	1	< 0.1%
0.05	1	< 0.1%
0.1	197	< 0.1%
0.12	1	< 0.1%
0.16	1	< 0.1%
0.2	185	< 0.1%
0.3	159	< 0.1%

Value	Count	Frequency (%)
892.3	1	< 0.1%
366.6	1	< 0.1%
180.3	1	< 0.1%
177.7	1	< 0.1%
153	1	< 0.1%
152.7	1	< 0.1%
150.7	1	< 0.1%
148	1	< 0.1%
146.1	1	< 0.1%
141.8	1	< 0.1%

total_acc

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: open_acc).

HIGH CORRELATION (This variable has a high correlation with 1 fields: open_acc).

HIGH CORRELATION (This variable has a high correlation with 1 fields: open_acc).

HIGH CORRELATION (This variable has a high correlation with 1 fields: open_acc).

Distinct

112

Distinct (%)

< 0.1%

Missing

29

Missing (%)

< 0.1%

Infinite

0

Infinite (%)

0.0%

Mean

25.06443027

Minimum

1

Maximum

156

Zeros

0

Zeros (%)

0.0%

Negative

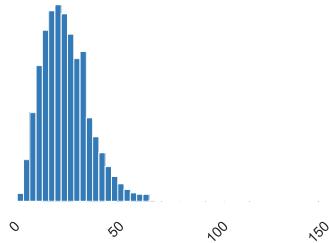
0

Negative (%)

0.0%

Memory size

3.6 MiB



Quantile statistics

Minimum

1

5-th percentile

9

Q1

17

median

23

Q3

32

95-th percentile

47

Maximum

156

Range

155

Interquartile range (IQR)

15

Descriptive statistics

Standard deviation

11.60014095

Coefficient of variation (CV)

0.4628128716

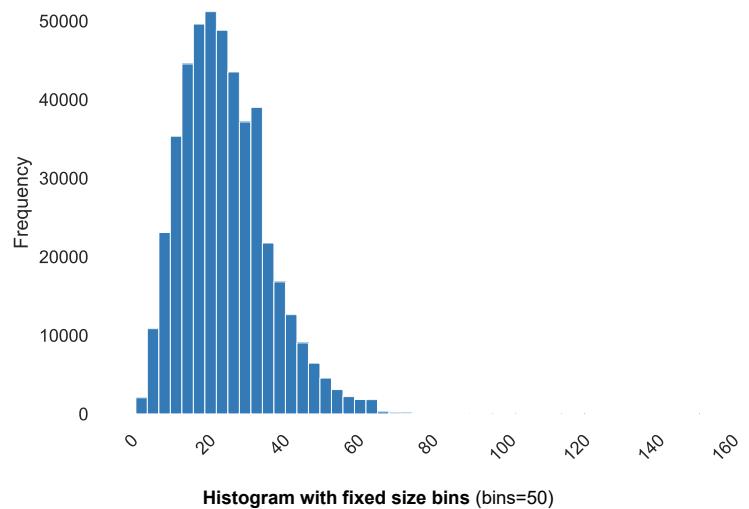
Kurtosis

0.9340984925

Mean

25.06443027

Median Absolute Deviation (MAD)	7
Skewness	0.8130610564
Sum	11686441
Variance	134.56327
Monotonicity	Not monotonic



Value	Count	Frequency (%)
21	17087	3.7%
20	17073	3.7%
22	17028	3.7%
23	16719	3.6%
19	16709	3.6%
18	16490	3.5%
17	16410	3.5%
24	16387	3.5%
25	15715	3.4%
16	15654	3.4%
Other values (102)	300984	64.5%

Value	Count	Frequency (%)
1	21	< 0.1%
2	62	< 0.1%
3	472	0.1%
4	1504	0.3%
5	2387	0.5%
6	3612	0.8%
7	4850	1.0%
8	6280	1.3%
9	7632	1.6%
10	9177	2.0%

Value	Count	Frequency (%)
156	1	< 0.1%
150	1	< 0.1%
124	1	< 0.1%
121	1	< 0.1%
119	1	< 0.1%
118	1	< 0.1%
117	1	< 0.1%
116	1	< 0.1%
106	1	< 0.1%
105	4	< 0.1%

initial_list_status

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: policy_code, application_type)

HIGH CORRELATION (This variable has a high correlation with 4 fields: Unnamed: 0, id, member_id, issue_d)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

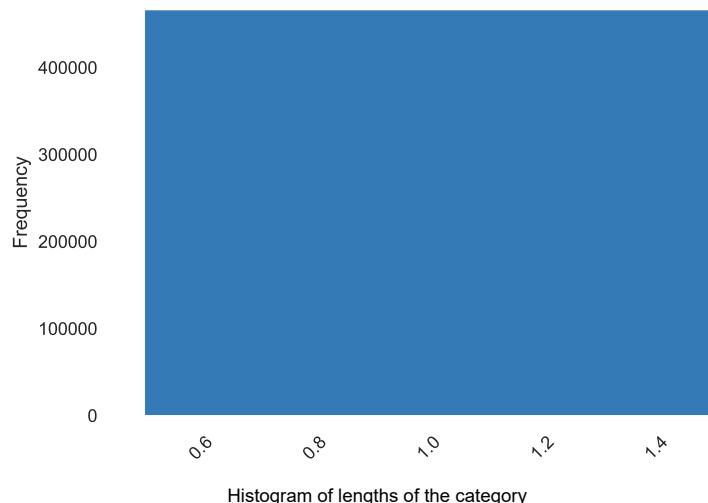
Sample

1st row	f
2nd row	f
3rd row	f
4th row	f
5th row	f

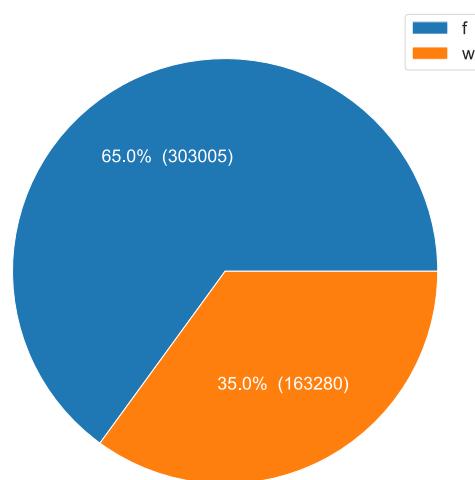
Common Values

Value	Count	Frequency (%)
f	303005	65.0%
w	163280	35.0%

Length



Pie chart



Value	Count	Frequency (%)
f	303005	65.0%
w	163280	35.0%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

out_prncp

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 3 fields: id, member_id, out_prncp_inv)

HIGH CORRELATION (This variable has a high correlation with 4 fields: loan_amnt, funded_amnt, funded_amnt_inv, out_prncp_inv)

HIGH CORRELATION (This variable has a high correlation with 1 fields: out_prncp_inv)

HIGH CORRELATION (This variable has a high correlation with 12 fields: Unnamed: 0, id, member_id, loan_amnt, funded_amnt, funded_amnt_inv, term, installment_loan_status, out_prncp_inv, total_rec_int, last_pymnt_d)

ZEROS

Distinct 135665

Distinct (%) 29.1%

Missing 0

Missing (%) 0.0%

Infinite 0

Infinite (%) 0.0%

Mean 4410.062342

Minimum 0

Maximum 32160.38

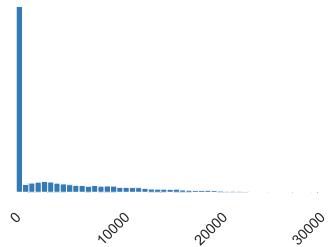
Zeros 229999

Zeros (%) 49.3%

Negative 0

Negative (%) 0.0%

Memory size 3.6 MiB



Quantile statistics

Minimum 0

5-th percentile 0

Q1 0

median 441.47

Q3 7341.65

95-th percentile 18365.82

Maximum 32160.38

Range 32160.38

Interquartile range (IQR) 7341.65

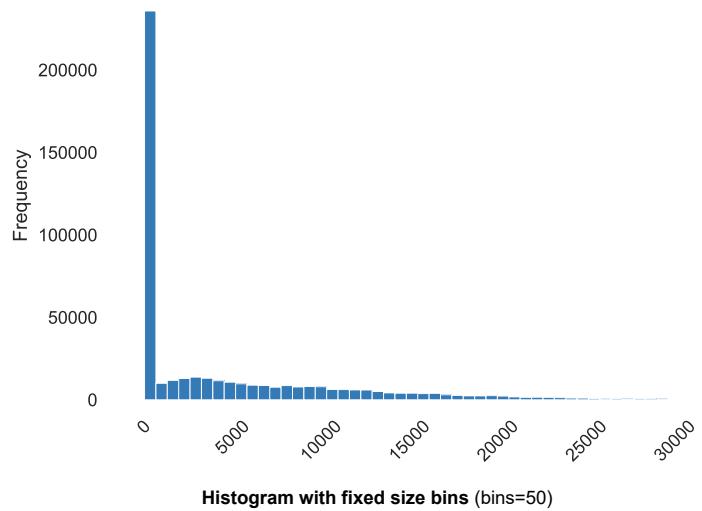
Descriptive statistics

Standard deviation 6355.078769

Coefficient of variation (CV) 1.441040574

Kurtosis 1.957280214

Mean	4410.062342
Median Absolute Deviation (MAD)	441.47
Skewness	1.590188568
Sum	2056345919
Variance	40387026.16
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	229999	49.3%
5684.61	80	< 0.1%
6563.61	79	< 0.1%
5617.88	78	< 0.1%
6110.65	76	< 0.1%
5709.44	75	< 0.1%
5131.53	71	< 0.1%
6227.69	71	< 0.1%
4519.91	70	< 0.1%
6821.49	68	< 0.1%
Other values (135655)	235618	50.5%

Value	Count	Frequency (%)
0	229999	49.3%
0.01	1	< 0.1%
0.03	1	< 0.1%
0.17	1	< 0.1%
0.21	1	< 0.1%
0.25	1	< 0.1%
0.38	1	< 0.1%
0.45	1	< 0.1%
0.62	1	< 0.1%
0.72	1	< 0.1%

Value	Count	Frequency (%)
32160.38	1	< 0.1%
31898.77	1	< 0.1%
31769.85	1	< 0.1%
31681.54	2	< 0.1%
31633.26	1	< 0.1%
31568.28	1	< 0.1%
31553.95	1	< 0.1%
31495.32	1	< 0.1%
31481.45	1	< 0.1%
31379.35	1	< 0.1%

out_prncp_inv

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 3 fields: id, member_id, out_prncp)

HIGH CORRELATION (This variable has a high correlation with 4 fields: loan_amnt, funded_amnt, funded_amnt_inv, out_prncp)

HIGH CORRELATION (This variable has a high correlation with 1 fields: out_prncp)

HIGH CORRELATION (This variable has a high correlation with 12 fields: Unnamed: 0, id, member_id, loan_amnt, funded_amnt, funded_amnt_inv, term, installment_loan_status, out_prncp, total_rec_int, last_pymnt_d)

ZEROS

Distinct	141189
----------	--------

Distinct (%)	30.3%
--------------	-------

Missing	0
---------	---

Missing (%)	0.0%
-------------	------

Infinite	0
----------	---

Infinite (%)	0.0%
--------------	------

Mean	4408.452258
------	-------------

Minimum	0
---------	---

Maximum	32160.38
---------	----------

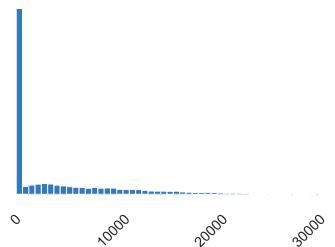
Zeros	229999
-------	--------

Zeros (%)	49.3%
-----------	-------

Negative	0
----------	---

Negative (%)	0.0%
--------------	------

Memory size	3.6 MiB
-------------	---------



Quantile statistics

Minimum	0
---------	---

5-th percentile	0
-----------------	---

Q1	0
----	---

median	441.38
--------	--------

Q3	7338.39
----	---------

95-th percentile	18353.936
------------------	-----------

Maximum	32160.38
---------	----------

Range	32160.38
-------	----------

Interquartile range (IQR)	7338.39
---------------------------	---------

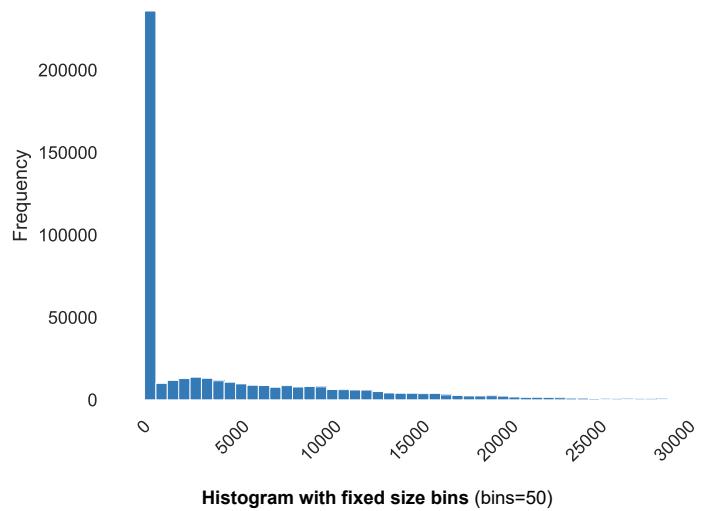
Descriptive statistics

Standard deviation	6353.198001
--------------------	-------------

Coefficient of variation (CV)	1.441140253
-------------------------------	-------------

Kurtosis	1.958572991
----------	-------------

Mean	4408.452258
Median Absolute Deviation (MAD)	441.38
Skewness	1.590502706
Sum	2055595161
Variance	40363124.84
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	229999	49.3%
6563.61	78	< 0.1%
5684.61	77	< 0.1%
5617.88	76	< 0.1%
6110.65	75	< 0.1%
5709.44	71	< 0.1%
6227.69	69	< 0.1%
5131.53	68	< 0.1%
4519.91	68	< 0.1%
6297.76	67	< 0.1%
Other values (141179)	235637	50.5%

Value	Count	Frequency (%)
0	229999	49.3%
0.01	1	< 0.1%
0.03	1	< 0.1%
0.17	1	< 0.1%
0.21	1	< 0.1%
0.25	1	< 0.1%
0.38	1	< 0.1%
0.45	1	< 0.1%
0.62	1	< 0.1%
0.72	1	< 0.1%

Value	Count	Frequency (%)
32160.38	1	< 0.1%
31898.77	1	< 0.1%
31769.85	1	< 0.1%
31681.54	2	< 0.1%
31633.26	1	< 0.1%
31568.28	1	< 0.1%
31531.42	1	< 0.1%
31495.32	1	< 0.1%
31481.45	1	< 0.1%
31379.35	1	< 0.1%

total_pymnt

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

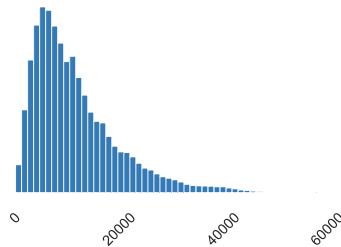
CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt).

HIGH CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt).

HIGH CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt_inv, total_rec_prncp, last_pymnt_amnt).

HIGH CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt_inv, total_rec_prncp, total_rec_int, last_pymnt_amnt).

Distinct	351609
Distinct (%)	75.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	11540.68622
Minimum	0
Maximum	57777.57987
Zeros	87
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



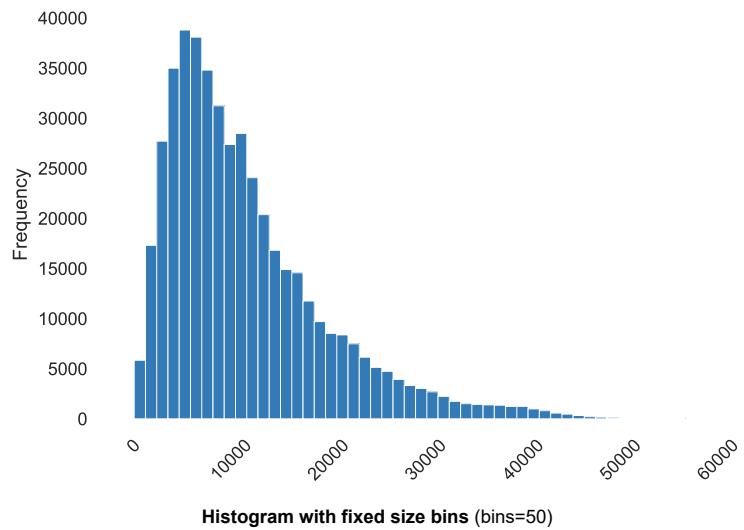
Quantile statistics

Minimum	0
5-th percentile	2319.472
Q1	5552.125349
median	9419.250943
Q3	15308.15846
95-th percentile	28312.938
Maximum	57777.57987
Range	57777.57987
Interquartile range (IQR)	9756.033111

Descriptive statistics

Standard deviation	8265.627112
Coefficient of variation (CV)	0.7162162592

Kurtosis	2.200556278
Mean	11540.68622
Median Absolute Deviation (MAD)	4468.929057
Skewness	1.406403305
Sum	5381248874
Variance	68320591.56
Monotonicity	Not monotonic



Value	Count	Frequency (%)
6743.73	117	< 0.1%
0	87	< 0.1%
5564.78	78	< 0.1%
5497.46	75	< 0.1%
5419.43	75	< 0.1%
5619.69	75	< 0.1%
6219.46	69	< 0.1%
7074.9	68	< 0.1%
5892.12	68	< 0.1%
6280.83	66	< 0.1%
Other values (351599)	465507	99.8%

Value	Count	Frequency (%)
0	87	< 0.1%
33.73	1	< 0.1%
34.88	1	< 0.1%
35.69	1	< 0.1%
35.71	1	< 0.1%
44.92	2	< 0.1%
44.96	1	< 0.1%
51.02	1	< 0.1%
57.18	1	< 0.1%
57.3	1	< 0.1%

Value	Count	Frequency (%)
57777.57987	1	< 0.1%
57060.13991	1	< 0.1%
56809.05163	1	< 0.1%
56087.90992	1	< 0.1%
55982.6499	1	< 0.1%
55906.9499	1	< 0.1%
55876.07267	1	< 0.1%
55759.48994	1	< 0.1%
55758.22208	1	< 0.1%
55444.75006	1	< 0.1%

total_pymnt_inv

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

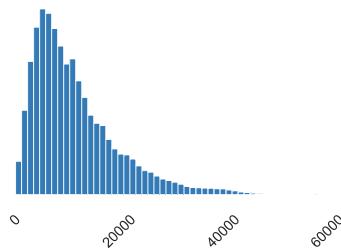
CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_rec_prncp, total_rec_int, last_pymnt_amnt)

HIGH CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_rec_prncp, total_rec_int, last_pymnt_amnt)

HIGH CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_rec_prncp, last_pymnt_amnt)

HIGH CORRELATION (This variable has a high correlation with 8 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_rec_prncp, total_rec_int, last_pymnt_amnt)

Distinct	347659
Distinct (%)	74.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	11469.89275
Minimum	0
Maximum	57777.58
Zeros	365
Zeros (%)	0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



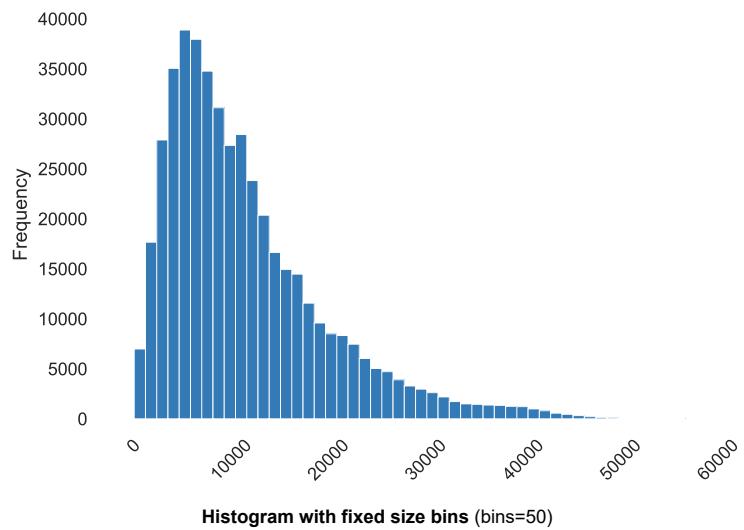
Quantile statistics

Minimum	0
5-th percentile	2244.66
Q1	5499.25
median	9355.43
Q3	15231.31
95-th percentile	28191.716
Maximum	57777.58
Range	57777.58
Interquartile range (IQR)	9732.06

Descriptive statistics

Standard deviation	8254.157579
Coefficient of variation (CV)	0.7196368581

Kurtosis	2.198647345
Mean	11469.89275
Median Absolute Deviation (MAD)	4456.28
Skewness	1.404658462
Sum	5348238939
Variance	68131117.33
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	365	0.1%
6743.73	113	< 0.1%
5564.78	77	< 0.1%
5419.43	73	< 0.1%
5497.46	72	< 0.1%
5619.69	71	< 0.1%
7074.9	66	< 0.1%
6219.46	66	< 0.1%
5892.12	65	< 0.1%
6280.83	64	< 0.1%
Other values (347649)	465253	99.8%

Value	Count	Frequency (%)
0	365	0.1%
0.49	1	< 0.1%
0.5	1	< 0.1%
0.92	1	< 0.1%
11.21	1	< 0.1%
12.65	1	< 0.1%
16.82	1	< 0.1%
18.97	1	< 0.1%
20.98	1	< 0.1%
21.6	1	< 0.1%

Value	Count	Frequency (%)
57777.58	1	< 0.1%
57019.38	1	< 0.1%
56475.05	1	< 0.1%
56047.85	1	< 0.1%
55982.65	1	< 0.1%
55876.07	1	< 0.1%
55867.02	1	< 0.1%
55759.49	1	< 0.1%
55145.01	1	< 0.1%
55061	1	< 0.1%

total_rec_prncp

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, last_pymnt_amnt)

HIGH CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, last_pymnt_amnt)

HIGH CORRELATION (This variable has a high correlation with 3 fields: total_pymnt, total_pymnt_inv, last_pymnt_amnt)

HIGH CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv, last_pymnt_amnt)

Distinct 172713

Distinct (%) 37.0%

Missing 0

Missing (%) 0.0%

Infinite 0

Infinite (%) 0.0%

Mean 8866.014657

Minimum 0

Maximum 35000.03

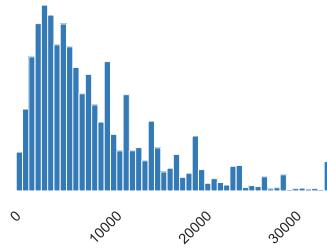
Zeros 399

Zeros (%) 0.1%

Negative 0

Negative (%) 0.0%

Memory size 3.6 MiB



Quantile statistics

Minimum 0

5-th percentile 1491.892

Q1 3708.56

median 6817.76

Q3 12000

95-th percentile 24000

Maximum 35000.03

Range 35000.03

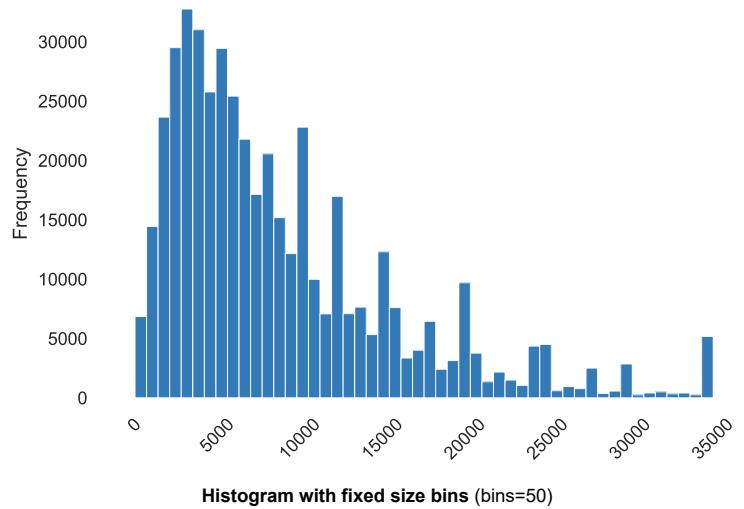
Interquartile range (IQR) 8291.44

Descriptive statistics

Standard deviation 7031.687997

Coefficient of variation (CV) 0.7931058395

Kurtosis	2.041532541
Mean	8866.014657
Median Absolute Deviation (MAD)	3610.92
Skewness	1.445591068
Sum	4134089644
Variance	49444636.09
Monotonicity	Not monotonic



Value	Count	Frequency (%)
10000	13267	2.8%
12000	10469	2.2%
15000	8552	1.8%
20000	7673	1.6%
8000	6522	1.4%
6000	6496	1.4%
5000	5968	1.3%
35000	4887	1.0%
16000	4352	0.9%
18000	4047	0.9%
Other values (172703)	394052	84.5%

Value	Count	Frequency (%)
0	399	0.1%
18.8	1	< 0.1%
19.31	1	< 0.1%
21.21	1	< 0.1%
21.34	1	< 0.1%
21.6	1	< 0.1%
21.63	1	< 0.1%
21.93	1	< 0.1%
22.24	1	< 0.1%
22.36	1	< 0.1%

Value	Count	Frequency (%)
35000.03	1	< 0.1%
35000.02	5	< 0.1%
35000.01	29	< 0.1%
35000	4887	1.0%
34999.99	68	< 0.1%
34999.98	8	< 0.1%
34999.97	11	< 0.1%
34999.96	10	< 0.1%
34999.95	8	< 0.1%
34999.92	1	< 0.1%

total_rec_int

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 6 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv)

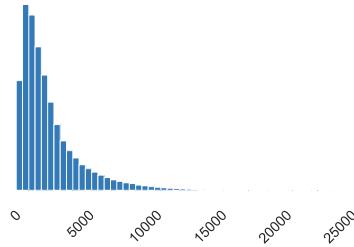
HIGH CORRELATION (This variable has a high correlation with 6 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment, total_pymnt, total_pymnt_inv)

total_pymnt_inv)

HIGH CORRELATION (This variable has a high correlation with 4 fields: loan_amnt, funded_amnt, funded_amnt_inv, installment)

HIGH CORRELATION (This variable has a high correlation with 10 fields: loan_amnt, funded_amnt, funded_amnt_inv, term, int_rate, installment, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv)

Distinct	270249
Distinct (%)	58.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2588.677225
Minimum	0
Maximum	24205.62
Zeros	379
Zeros (%)	0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



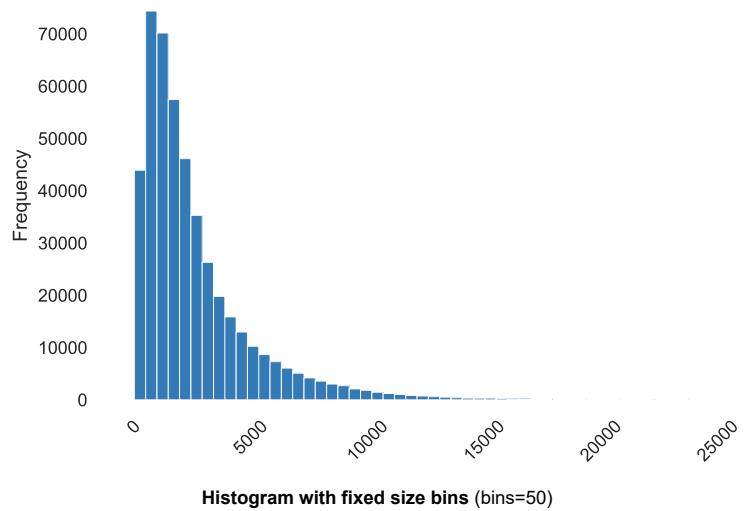
Quantile statistics

Minimum	0
5-th percentile	317.76
Q1	957.28
median	1818.88
Q3	3304.53
95-th percentile	7640.926
Maximum	24205.62
Range	24205.62
Interquartile range (IQR)	2347.25

Descriptive statistics

Standard deviation	2483.809661
Coefficient of variation (CV)	0.9594899038
Kurtosis	7.158621674

Mean	2588.677225
Median Absolute Deviation (MAD)	1027.07
Skewness	2.26827618
Sum	1207061360
Variance	6169310.434
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	379	0.1%
1249.39	78	< 0.1%
1329.13	75	< 0.1%
1037.31	74	< 0.1%
1350.99	72	< 0.1%
1263.64	71	< 0.1%
1499.26	68	< 0.1%
1542.27	65	< 0.1%
1437.51	65	< 0.1%
1716.03	65	< 0.1%
Other values (270239)	465273	99.8%

Value	Count	Frequency (%)
0	379	0.1%
1	1	< 0.1%
1.16	1	< 0.1%
1.34	1	< 0.1%
1.67	1	< 0.1%
3.18	1	< 0.1%
3.54	1	< 0.1%
3.74	2	< 0.1%
3.94	1	< 0.1%
5	1	< 0.1%

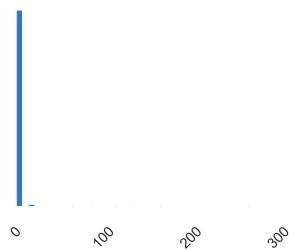
Value	Count	Frequency (%)
24205.62	1	< 0.1%
23450.38	1	< 0.1%
23172.31	1	< 0.1%
23171.65	1	< 0.1%
23062.45	1	< 0.1%
22930.84	1	< 0.1%
22891.95	1	< 0.1%
22886.89	1	< 0.1%
22777.58	1	< 0.1%
22745.22	1	< 0.1%

total_rec_late_fee

Real number ($\mathbb{R}_{\geq 0}$)

ZEROS

Distinct	5808
Distinct (%)	1.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.6501291808
Minimum	0
Maximum	358.68
Zeros	455607
Zeros (%)	97.7%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



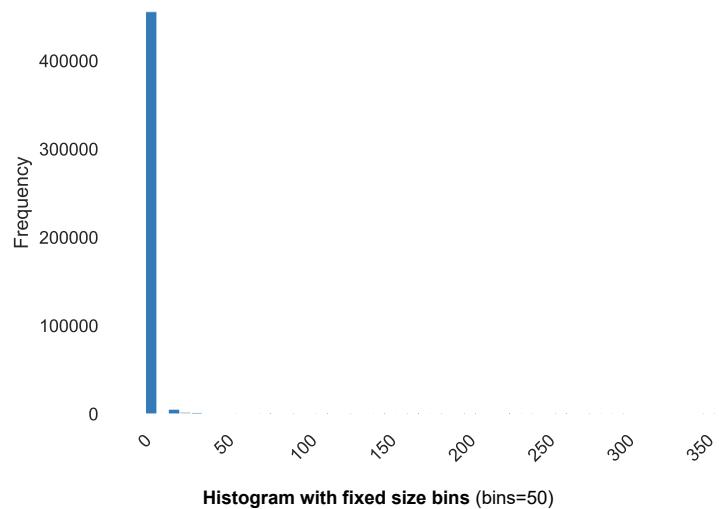
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	358.68
Range	358.68
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	5.265729688
Coefficient of variation (CV)	8.099512903
Kurtosis	329.0750629
Mean	0.6501291808
Median Absolute Deviation (MAD)	0
Skewness	13.84349742
Sum	303145.4851

Variance	27.72790915
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	455607	97.7%
15	2223	0.5%
30	285	0.1%
45	50	< 0.1%
16.84	15	< 0.1%
17.41	14	< 0.1%
15.00000001	14	< 0.1%
16.64	14	< 0.1%
17.01	13	< 0.1%
60	13	< 0.1%
Other values (5798)	8037	1.7%

Value	Count	Frequency (%)
0	455607	97.7%
0.01	6	< 0.1%
0.060799751	1	< 0.1%
0.073787104	1	< 0.1%
0.081837685	1	< 0.1%
0.101704562	1	< 0.1%
0.135765627	1	< 0.1%
0.180082904	1	< 0.1%
0.18477362	1	< 0.1%
0.2	2	< 0.1%

Value	Count	Frequency (%)
358.68	1	< 0.1%
294.68	1	< 0.1%
286.7475657	1	< 0.1%
268.55	1	< 0.1%
252.8	1	< 0.1%
242.55	1	< 0.1%
229.75	1	< 0.1%
213.3	1	< 0.1%
208.81953	1	< 0.1%
199.16	1	< 0.1%

recoveries

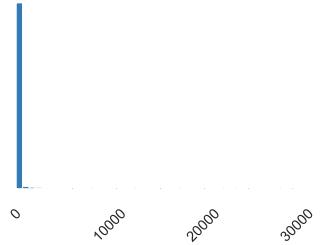
Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: collection_recovery_fee).
HIGH CORRELATION (This variable has a high correlation with 1 fields: collection_recovery_fee).
HIGH CORRELATION (This variable has a high correlation with 1 fields: collection_recovery_fee).
HIGH CORRELATION (This variable has a high correlation with 1 fields: collection_recovery_fee).

ZEROS

Distinct	22773
Distinct (%)	4.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	85.34421125
Minimum	0
Maximum	33520.27
Zeros	441914
Zeros (%)	94.8%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



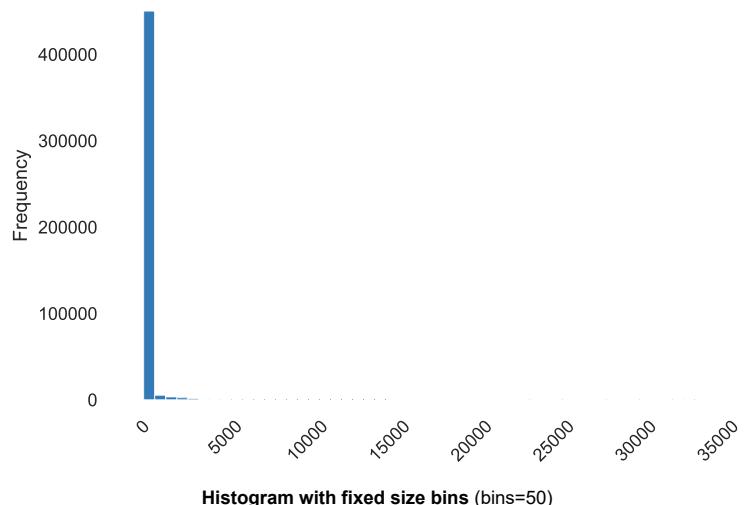
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	16.864
Maximum	33520.27
Range	33520.27
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	552.2160836
Coefficient of variation (CV)	6.470457404
Kurtosis	330.0485701
Mean	85.34421125

Median Absolute Deviation (MAD)	0
Skewness	13.14994237
Sum	39794725.54
Variance	304942.603
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	441914	94.8%
14.28	9	< 0.1%
11.7	9	< 0.1%
200	9	< 0.1%
10	9	< 0.1%
16.5	8	< 0.1%
16.8	8	< 0.1%
800	8	< 0.1%
10.4	8	< 0.1%
100	8	< 0.1%
Other values (22763)	24295	5.2%

Value	Count	Frequency (%)
0	441914	94.8%
0.25	1	< 0.1%
2.34	1	< 0.1%
3.6	1	< 0.1%
4.4	1	< 0.1%
6.3	1	< 0.1%
6.31	1	< 0.1%
6.87	1	< 0.1%
7.11	1	< 0.1%
8.19	1	< 0.1%

Value	Count	Frequency (%)
33520.27	1	< 0.1%
31900.52	1	< 0.1%
29623.35	1	< 0.1%
29282.07	1	< 0.1%
27750	1	< 0.1%
25000.29	1	< 0.1%
24862.1	1	< 0.1%
24833.68	1	< 0.1%
23184.33	1	< 0.1%
23035.92	1	< 0.1%

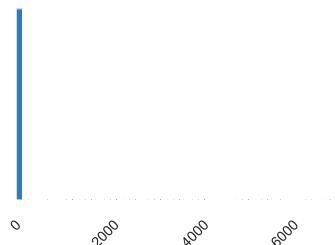
collection_recovery_fee

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: recoveries)
HIGH CORRELATION (This variable has a high correlation with 1 fields: recoveries)
HIGH CORRELATION (This variable has a high correlation with 1 fields: recoveries)
HIGH CORRELATION (This variable has a high correlation with 1 fields: recoveries)
SKEWED
ZEROS

Distinct	20275
Distinct (%)	4.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	8.961533902
Minimum	0
Maximum	7002.19
Zeros	443079
Zeros (%)	95.0%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



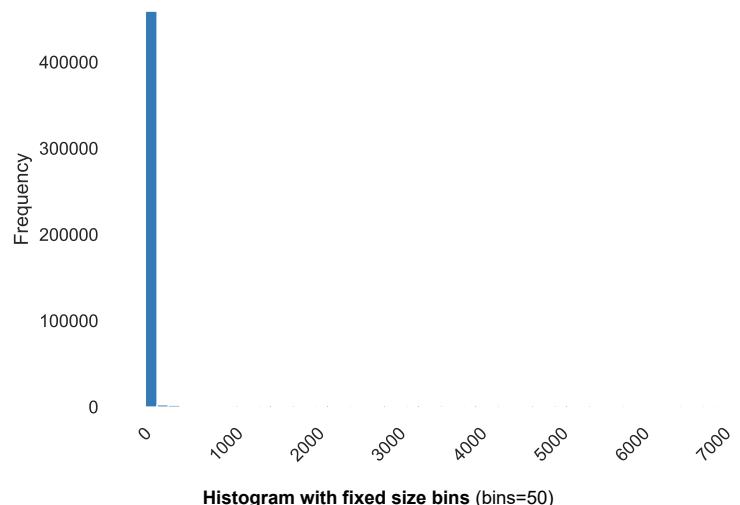
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	7002.19
Range	7002.19
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	85.491437
Coefficient of variation (CV)	9.539821859
Kurtosis	942.9795228

Mean	8.961533902
Median Absolute Deviation (MAD)	0
Skewness	22.23609553
Sum	4178628.835
Variance	7308.785801
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	443079	95.0%
2	18	< 0.1%
1.8	13	< 0.1%
1.88	12	< 0.1%
1.2	12	< 0.1%
1.55	12	< 0.1%
3.2	11	< 0.1%
4.52	11	< 0.1%
1.6	11	< 0.1%
2.52	11	< 0.1%
Other values (20265)	23095	5.0%

Value	Count	Frequency (%)
0	443079	95.0%
0.036	1	< 0.1%
0.045	1	< 0.1%
0.063	1	< 0.1%
0.071099997	1	< 0.1%
0.074500001	1	< 0.1%
0.088499999	1	< 0.1%
0.1181	1	< 0.1%
0.134799995	1	< 0.1%
0.1393	1	< 0.1%

Value	Count	Frequency (%)
7002.19	1	< 0.1%
6972.59	1	< 0.1%
6543.04	1	< 0.1%
5774.8	1	< 0.1%
5694.0936	1	< 0.1%
5602.72	1	< 0.1%
5569.92	1	< 0.1%
5216.74	1	< 0.1%
5192.99	1	< 0.1%
5036.01	1	< 0.1%

last_pymnt_d

Categorical

HIGH

CARDINALITY

HIGH CORRELATION (This variable has a high correlation with 3 fields: policy_code, application_type, next_pymnt_d)HIGH CORRELATION (This variable has a high correlation with 9 fields: Unnamed: 0, id, member_id, issue_d, loan_status, mths_since_last_record, out_prncp, out_prncp_inv, next_pymnt_d)

Distinct	98
Distinct (%)	< 0.1%
Missing	376
Missing (%)	0.1%
Memory size	3.6 MiB

Length

Max length	6
Median length	6
Mean length	6
Min length	6

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

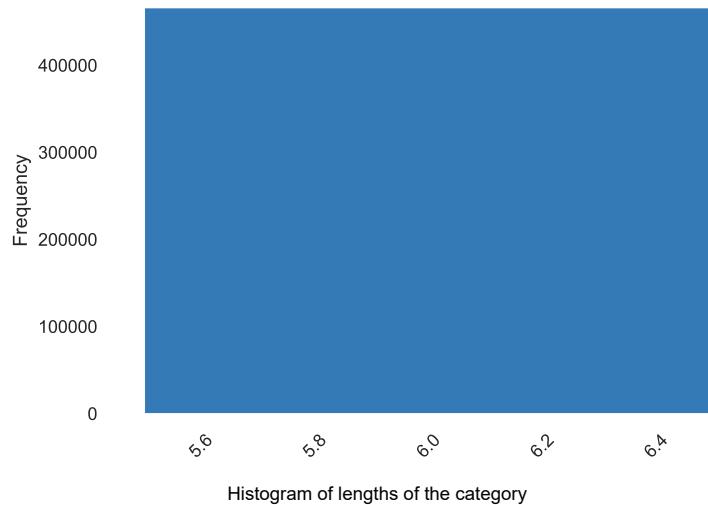
Sample

1st row	Jan-15
2nd row	Apr-13
3rd row	Jun-14
4th row	Jan-15
5th row	Jan-16

Common Values

Value	Count	Frequency (%)
Jan-16	179620	38.5%
Dec-15	62000	13.3%
Jul-15	11703	2.5%
Oct-15	11248	2.4%
Sep-15	10204	2.2%
Mar-15	10087	2.2%
Jun-15	10078	2.2%
Apr-15	10051	2.2%
Aug-15	9939	2.1%
May-15	9618	2.1%
Other values (88)	141361	30.3%

Length



Value	Count	Frequency (%)
jan-16	179620	38.6%
dec-15	62000	13.3%
jul-15	11703	2.5%
oct-15	11248	2.4%
sep-15	10204	2.2%
mar-15	10087	2.2%
jun-15	10078	2.2%
apr-15	10051	2.2%
aug-15	9939	2.1%
may-15	9618	2.1%
Other values (88)	141361	30.3%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

last_pymnt_amnt

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

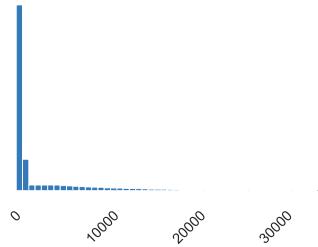
CORRELATION (This variable has a high correlation with 4 fields: installment, total_pymnt, total_pymnt_inv, total_rec_prncp)

HIGH CORRELATION (This variable has a high correlation with 3 fields: total_pymnt, total_pymnt_inv, total_rec_prncp)

HIGH CORRELATION (This variable has a high correlation with 3 fields: total_pymnt, total_pymnt_inv, total_rec_prncp)

HIGH CORRELATION (This variable has a high correlation with 7 fields: loan_amnt, funded_amnt, funded_amnt_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, next_pymnt_d)

Distinct	198194
Distinct (%)	42.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3123.913796
Minimum	0
Maximum	36234.44
Zeros	390
Zeros (%)	0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



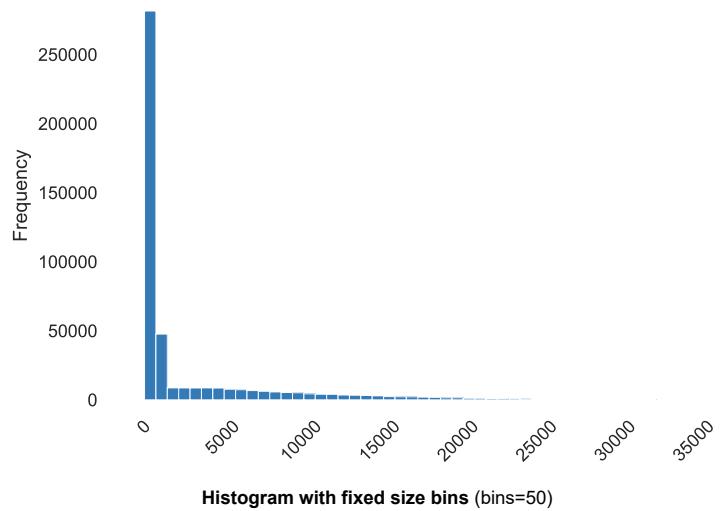
Quantile statistics

Minimum	0
5-th percentile	121.692
Q1	312.62
median	545.96
Q3	3187.51
95-th percentile	15735.668
Maximum	36234.44
Range	36234.44
Interquartile range (IQR)	2874.89

Descriptive statistics

Standard deviation	5554.737393
Coefficient of variation (CV)	1.778134019
Kurtosis	6.930578171
Mean	3123.913796

Median Absolute Deviation (MAD)	316.32
Skewness	2.560644376
Sum	1456634144
Variance	30855107.51
Monotonicity	Not monotonic



Value	Count	Frequency (%)
327.34	860	0.2%
336.9	828	0.2%
332.1	625	0.1%
491.01	579	0.1%
392.81	567	0.1%
334.49	543	0.1%
608.72	465	0.1%
654.68	446	0.1%
269.52	440	0.1%
317.54	436	0.1%
Other values (198184)	460496	98.8%

Value	Count	Frequency (%)
0	390	0.1%
0.01	19	< 0.1%
0.02	4	< 0.1%
0.03	7	< 0.1%
0.04	1	< 0.1%
0.05	1	< 0.1%
0.06	1	< 0.1%
0.07	1	< 0.1%
0.08	1	< 0.1%
0.09	4	< 0.1%

Value	Count	Frequency (%)
36234.44	1	< 0.1%
36170.14	1	< 0.1%
36115.2	1	< 0.1%
36043.81	1	< 0.1%
36004.61	1	< 0.1%
35967.61	1	< 0.1%
35867.09	1	< 0.1%
35814.33	1	< 0.1%
35760.2	1	< 0.1%
35757.36	1	< 0.1%

next_pymnt_d

Categorical

HIGH

CARDINALITY

HIGH CORRELATION (This variable has a high correlation with 3 fields: last_pymnt_d, policy_code, application_type)HIGH CORRELATION (This variable has a high correlation with 7 fields: Unnamed: 0, issue_d, loan_status, ing_last_6mths, mths_since_last_record, last_pymnt_d, last_pymnt_amnt)

MISSING

Distinct	100
Distinct (%)	< 0.1%
Missing	227214
Missing (%)	48.7%
Memory size	3.6 MiB

Length

Max length	6
Median length	6
Mean length	6
Min length	6

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	3	
Unique (%)	< 0.1%	

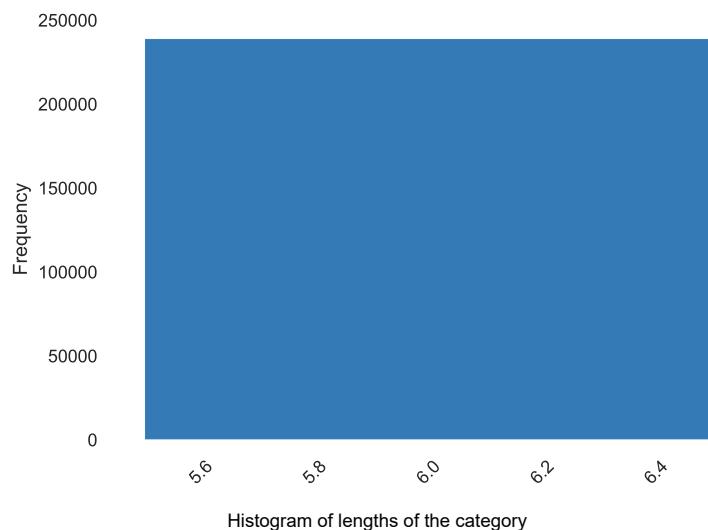
Sample

1st row	Feb-16
2nd row	Feb-16
3rd row	Jan-16
4th row	Jan-16
5th row	Jan-16

Common Values

Value	Count	Frequency (%)
Feb-16	208393	44.7%
Jan-16	27922	6.0%
Mar-11	107	< 0.1%
Apr-11	101	< 0.1%
Feb-11	91	< 0.1%
Jan-11	79	< 0.1%
May-11	77	< 0.1%
Dec-10	71	< 0.1%
Jun-11	66	< 0.1%
Sep-11	63	< 0.1%
Other values (90)	2101	0.5%
(Missing)	227214	48.7%

Length



Value	Count	Frequency (%)
feb-16	208393	87.2%
jan-16	27922	11.7%
mar-11	107	< 0.1%
apr-11	101	< 0.1%
feb-11	91	< 0.1%
jan-11	79	< 0.1%
may-11	77	< 0.1%
dec-10	71	< 0.1%
jun-11	66	< 0.1%
sep-11	63	< 0.1%
Other values (90)	2101	0.9%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

last_credit_pull_d

Categorical

HIGH

CARDINALITY

Distinct

103

Distinct (%)

< 0.1%

Missing

42

Missing (%)

< 0.1%

Memory size

3.6 MiB

Length

Max length	6
Median length	6
Mean length	6
Min length	6

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	4	?
Unique (%)	< 0.1%	

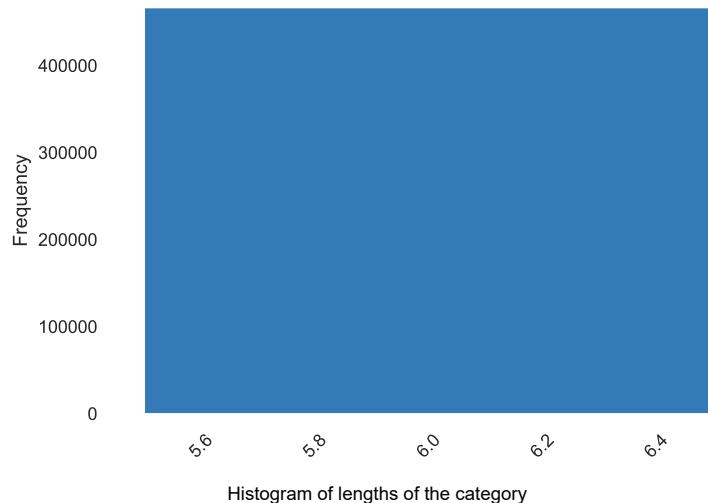
Sample

1st row	Jan-16
2nd row	Sep-13
3rd row	Jan-16
4th row	Jan-15
5th row	Jan-16

Common Values

Value	Count	Frequency (%)
Jan-16	327699	70.3%
Dec-15	13992	3.0%
Nov-15	8384	1.8%
Sep-15	8051	1.7%
Oct-15	7957	1.7%
Jul-15	7351	1.6%
Aug-15	6574	1.4%
Jun-15	6214	1.3%
Apr-15	5698	1.2%
Mar-15	5479	1.2%
Other values (93)	68844	14.8%

Length



Value	Count	Frequency (%)
jan-16	327699	70.3%
dec-15	13992	3.0%
nov-15	8384	1.8%
sep-15	8051	1.7%
oct-15	7957	1.7%
jul-15	7351	1.6%
aug-15	6574	1.4%
jun-15	6214	1.3%
apr-15	5698	1.2%
mar-15	5479	1.2%
Other values (93)	68844	14.8%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

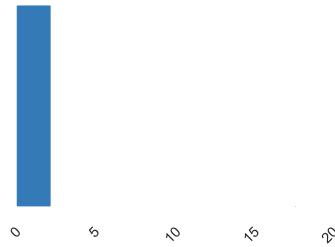
collections_12_mths_ex_med

Real number ($\mathbb{R}_{\geq 0}$)

SKEWED

ZEROS

Distinct	9
Distinct (%)	< 0.1%
Missing	145
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	0.009085253357
Minimum	0
Maximum	20
Zeros	462226
Zeros (%)	99.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



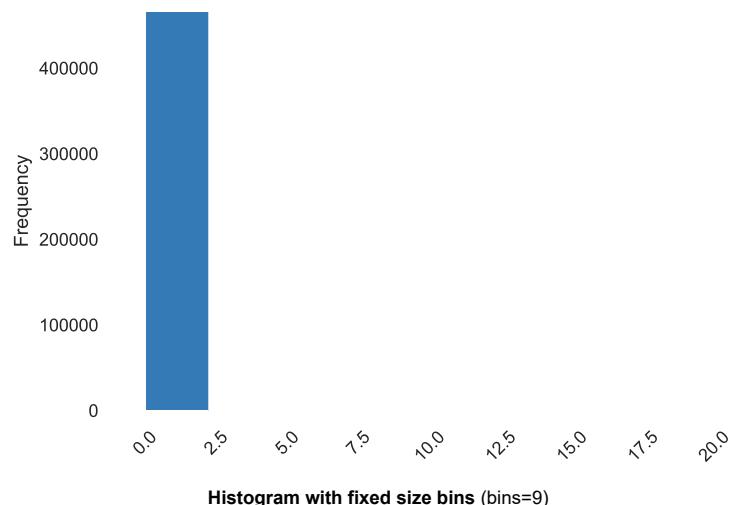
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	20
Range	20
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.1086483789
Coefficient of variation (CV)	11.95876159
Kurtosis	3650.010476
Mean	0.009085253357
Median Absolute Deviation (MAD)	0
Skewness	31.22030218

Sum	4235
Variance	0.01180447024
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	462226	99.1%
1	3665	0.8%
2	219	< 0.1%
3	19	< 0.1%
4	7	< 0.1%
6	1	< 0.1%
16	1	< 0.1%
20	1	< 0.1%
5	1	< 0.1%
(Missing)	145	< 0.1%

Value	Count	Frequency (%)
0	462226	99.1%
1	3665	0.8%
2	219	< 0.1%
3	19	< 0.1%
4	7	< 0.1%
5	1	< 0.1%
6	1	< 0.1%
16	1	< 0.1%
20	1	< 0.1%

Value	Count	Frequency (%)
20	1	< 0.1%
16	1	< 0.1%
6	1	< 0.1%
5	1	< 0.1%
4	7	< 0.1%
3	19	< 0.1%
2	219	< 0.1%
1	3665	0.8%
0	462226	99.1%

mths_since_last_major_derog

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: delinq_2yrs, mths_since_last_delinq)

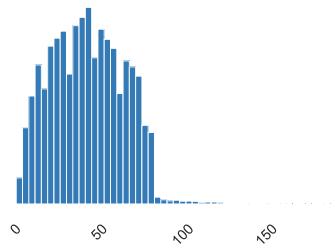
HIGH CORRELATION (This variable has a high correlation with 1 fields: mths_since_last_delinq)

HIGH CORRELATION (This variable has a high correlation with 1 fields: mths_since_last_delinq)

HIGH CORRELATION (This variable has a high correlation with 1 fields: mths_since_last_delinq)

MISSING

Distinct	162
Distinct (%)	0.2%
Missing	367311
Missing (%)	78.8%
Infinite	0
Infinite (%)	0.0%
Mean	42.85254713
Minimum	0
Maximum	188
Zeros	79
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



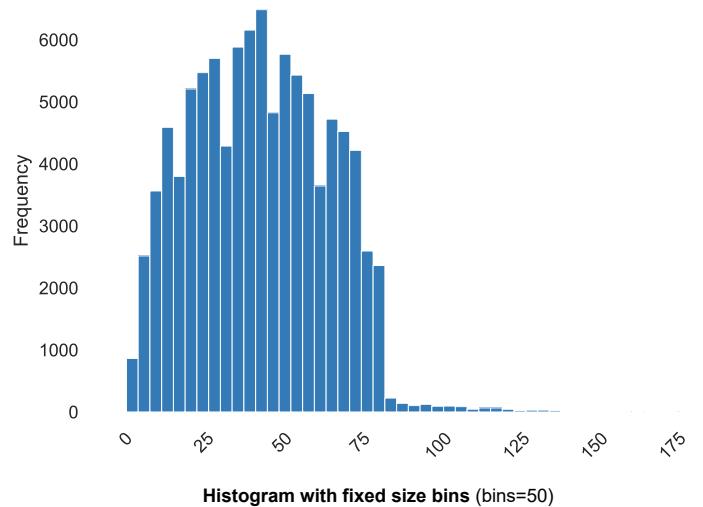
Quantile statistics

Minimum	0
5-th percentile	9
Q1	26
median	42
Q3	59
95-th percentile	77
Maximum	188
Range	188
Interquartile range (IQR)	33

Descriptive statistics

Standard deviation	21.6625913
Coefficient of variation (CV)	0.5055146717
Kurtosis	-0.1630539602
Mean	42.85254713

Median Absolute Deviation (MAD)	17
Skewness	0.2677133352
Sum	4241288
Variance	469.2678617
Monotonicity	Not monotonic



Value	Count	Frequency (%)
45	1672	0.4%
46	1670	0.4%
42	1656	0.4%
48	1637	0.4%
41	1592	0.3%
44	1583	0.3%
43	1575	0.3%
40	1567	0.3%
49	1541	0.3%
47	1517	0.3%
Other values (152)	82964	17.8%
(Missing)	367311	78.8%

Value	Count	Frequency (%)
0	79	< 0.1%
1	250	0.1%
2	245	0.1%
3	289	0.1%
4	429	0.1%
5	594	0.1%
6	704	0.2%
7	789	0.2%
8	855	0.2%
9	889	0.2%

Value	Count	Frequency (%)
188	1	< 0.1%
180	1	< 0.1%
170	1	< 0.1%
169	1	< 0.1%
165	1	< 0.1%
162	1	< 0.1%
159	1	< 0.1%
156	1	< 0.1%
154	2	< 0.1%
152	2	< 0.1%

policy_code

Categorical

CONSTANT

HTGH CORRELATION (This variable has a high correlation with 15 fields: grade, purpose, initial list status, addr state, last pymnt d, issue d, loan status, application type, sub grade, term, home ownership, emp length, pymnt plan, verification status, next pymnt d).

REJECTED

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

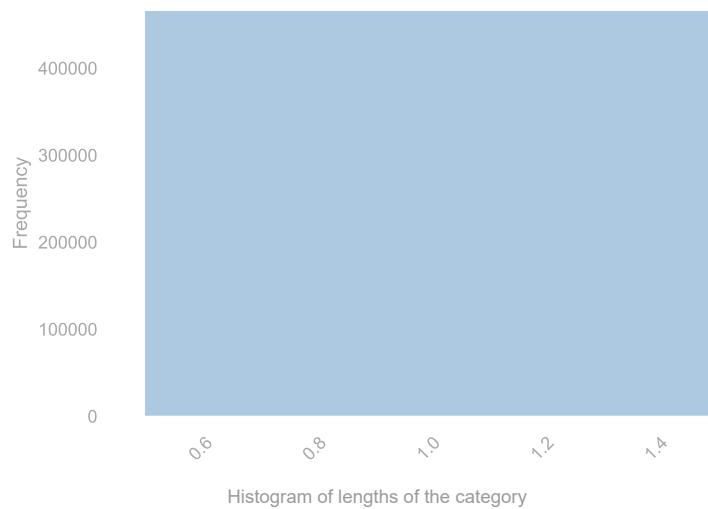
Sample

1st row	1
2nd row	1
3rd row	1
4th row	1
5th row	1

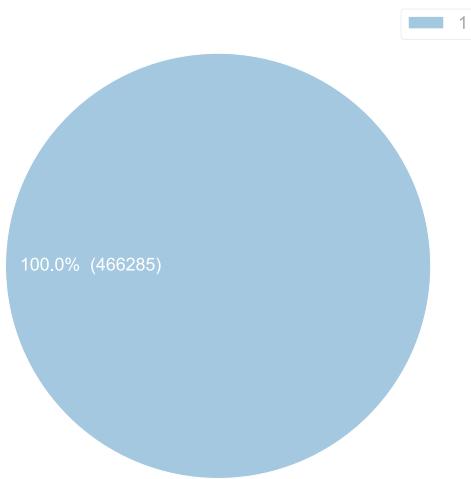
Common Values

Value	Count	Frequency (%)
1	466285	100.0%

Length



Pie chart



Value	Count	Frequency (%)
1	466285	100.0%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

application_type

Categorical

CONSTANT

HTGH CORRELATION (This variable has a high correlation with 15 fields: grade, purpose, initial list status, addr state, last pymnt d, issue d, loan status, policy code, sub grade, term, home ownership, emp length, pymnt plan, verification status, next pymnt d).

REJECTED

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.6 MiB

Length

Max length	10
Median length	10
Mean length	10
Min length	10

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

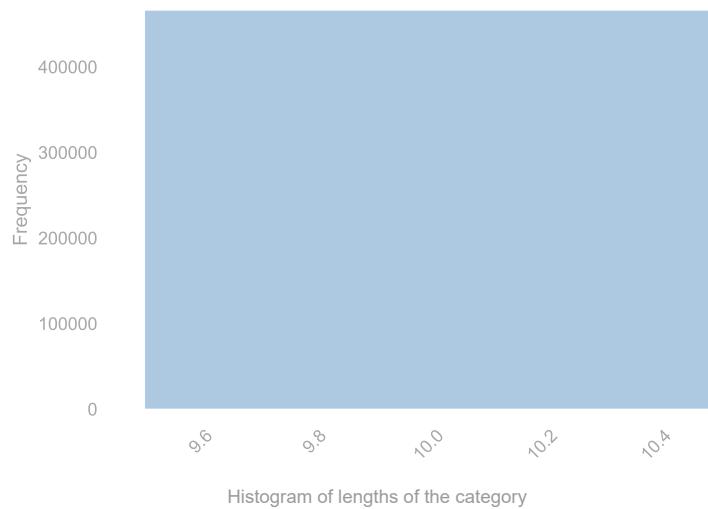
Sample

1st row	INDIVIDUAL
2nd row	INDIVIDUAL
3rd row	INDIVIDUAL
4th row	INDIVIDUAL
5th row	INDIVIDUAL

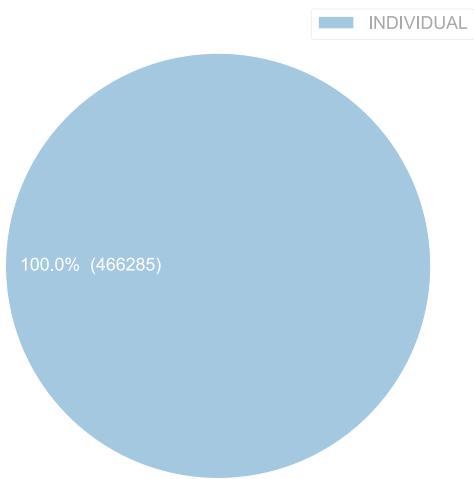
Common Values

Value	Count	Frequency (%)
INDIVIDUAL	466285	100.0%

Length



Pie chart



Value	Count	Frequency (%)
individual	466285	100.0%

Most occurring characters

Value

Count Frequency (%)

No values found.

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

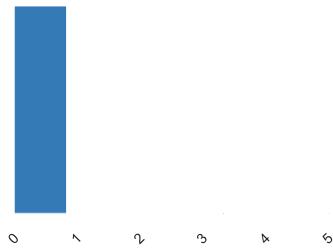
Most frequent character per block

annual_inc_joint		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

dti_joint		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

verification_status_joint		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

acc_now_delinq		
Real number ($\mathbb{R}_{\geq 0}$)		SKEWED
ZEROS		
Distinct	6	
Distinct (%)	< 0.1%	
Missing	29	
Missing (%)	< 0.1%	
Infinite	0	
Infinite (%)	0.0%	
Mean	0.004002093271	
Minimum	0	
Maximum	5	
Zeros	464524	
Zeros (%)	99.6%	
Negative	0	
Negative (%)	0.0%	
Memory size	3.6 MiB	

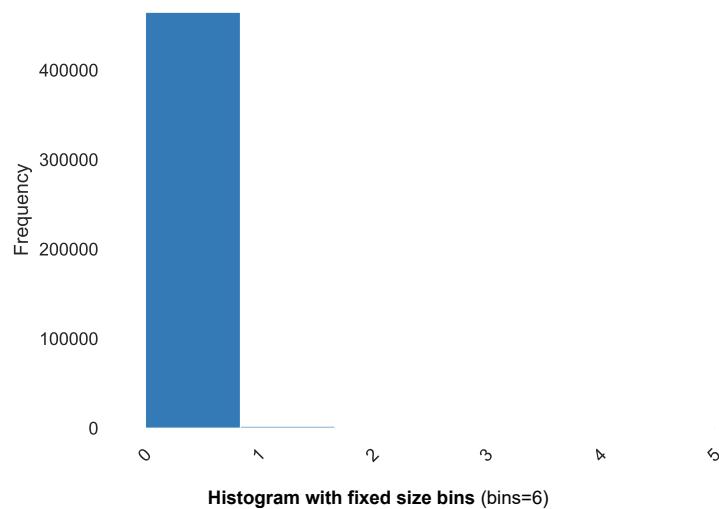


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	5
Range	5
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.06863679969
Coefficient of variation (CV)	17.15022491
Kurtosis	597.0909873
Mean	0.004002093271
Median Absolute Deviation (MAD)	0
Skewness	20.79829116
Sum	1866
Variance	0.004711010271
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	464524	99.6%
1	1623	0.3%
2	92	< 0.1%
3	11	< 0.1%
4	4	< 0.1%
5	2	< 0.1%
(Missing)	29	< 0.1%

Value	Count	Frequency (%)
0	464524	99.6%
1	1623	0.3%
2	92	< 0.1%
3	11	< 0.1%
4	4	< 0.1%
5	2	< 0.1%

Value	Count	Frequency (%)
5	2	< 0.1%
4	4	< 0.1%
3	11	< 0.1%
2	92	< 0.1%
1	1623	0.3%
0	464524	99.6%

tot_coll_amtReal number ($\mathbb{R}_{\geq 0}$)

MISSING

SKEWED

ZEROS

Distinct

6321

Distinct (%)

1.6%

Missing

70276

Missing (%)

15.1%

Infinite

0

Infinite (%)

0.0%

Mean

191.9135171

Minimum

0

Maximum

9152545

Zeros

346259

Zeros (%)

74.3%

Negative

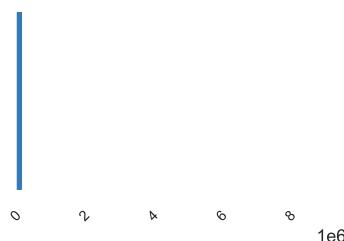
0

Negative (%)

0.0%

Memory size

3.6 MiB

**Quantile statistics****Minimum**

0

5-th percentile

0

Q1

0

median

0

Q3

0

95-th percentile

463

Maximum

9152545

Range

9152545

Interquartile range (IQR)

0

Descriptive statistics**Standard deviation**

14630.2146

Coefficient of variation (CV)

76.23337232

Kurtosis

386748.0772

Mean

191.9135171

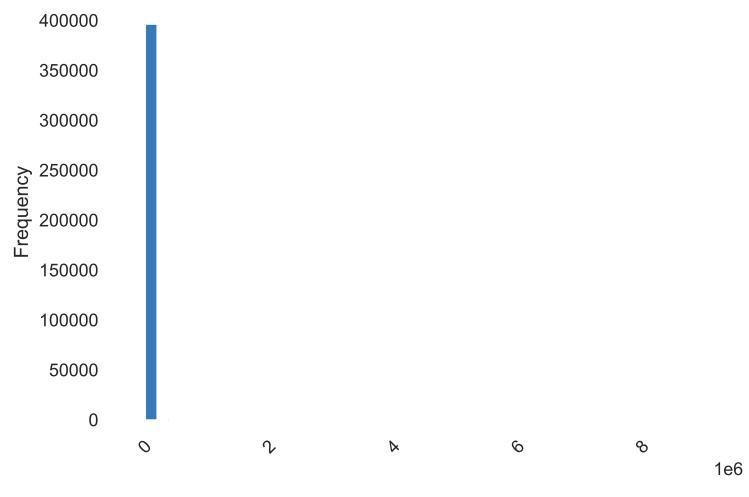
Median Absolute Deviation (MAD)

0

Skewness

618.2730112

Sum	75999480
Variance	214043179.4
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	346259	74.3%
50	837	0.2%
100	650	0.1%
75	515	0.1%
60	309	0.1%
200	309	0.1%
150	300	0.1%
70	281	0.1%
55	275	0.1%
80	268	0.1%
Other values (6311)	46006	9.9%
(Missing)	70276	15.1%

Value	Count	Frequency (%)
0	346259	74.3%
3	1	< 0.1%
5	1	< 0.1%
7	2	< 0.1%
9	1	< 0.1%
10	7	< 0.1%
12	1	< 0.1%
15	2	< 0.1%
17	1	< 0.1%
20	1	< 0.1%

Value	Count	Frequency (%)
9152545	1	< 0.1%
296368	1	< 0.1%
126297	1	< 0.1%
124825	1	< 0.1%
119428	1	< 0.1%
117800	1	< 0.1%
102752	1	< 0.1%
95806	1	< 0.1%
92221	1	< 0.1%
90116	1	< 0.1%

tot_cur_bal

Real number ($\mathbb{R}_{\geq 0}$)

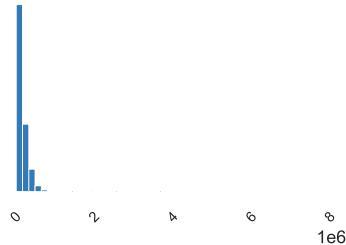
HIGH

CORRELATION (This variable has a high correlation with 1 fields: annual_inc)

HIGH CORRELATION (This variable has a high correlation with 1 fields: revol_bal)

MISSING

Distinct	220690
Distinct (%)	55.7%
Missing	70276
Missing (%)	15.1%
Infinite	0
Infinite (%)	0.0%
Mean	138801.7134
Minimum	0
Maximum	8000078
Zeros	78
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



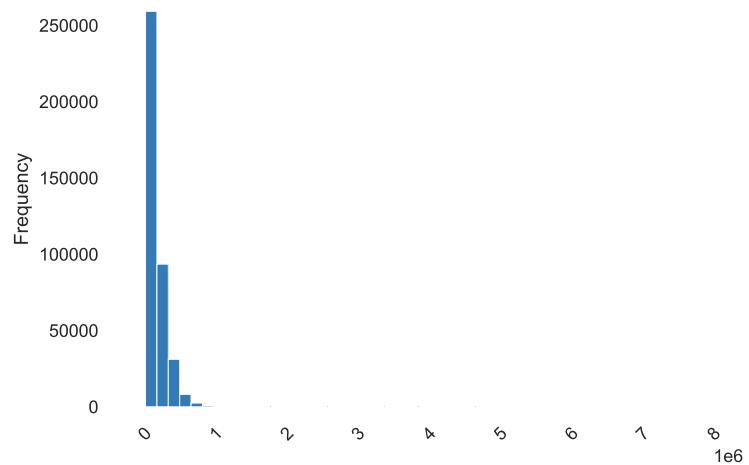
Quantile statistics

Minimum	0
5-th percentile	8725
Q1	28618
median	81539
Q3	208953
95-th percentile	419395.6
Maximum	8000078
Range	8000078
Interquartile range (IQR)	180335

Descriptive statistics

Standard deviation	152114.6635
Coefficient of variation (CV)	1.095913442
Kurtosis	39.77663669
Mean	138801.7134
Median Absolute Deviation (MAD)	65057

Skewness	3.047906078
Sum	$5.496672772 \times 10^{10}$
Variance	$2.313887085 \times 10^{10}$
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	78	< 0.1%
14186	15	< 0.1%
23254	15	< 0.1%
28674	14	< 0.1%
30491	14	< 0.1%
4875	13	< 0.1%
30652	13	< 0.1%
13818	13	< 0.1%
9134	12	< 0.1%
18583	12	< 0.1%
Other values (220680)	395810	84.9%
(Missing)	70276	15.1%

Value	Count	Frequency (%)
0	78	< 0.1%
1	1	< 0.1%
2	2	< 0.1%
5	2	< 0.1%
6	1	< 0.1%
9	1	< 0.1%
10	1	< 0.1%
11	3	< 0.1%
14	1	< 0.1%
15	3	< 0.1%

Value	Count	Frequency (%)
8000078	1	< 0.1%
4772549	1	< 0.1%
4026405	1	< 0.1%
3881449	1	< 0.1%
3840795	1	< 0.1%
3796811	1	< 0.1%
3610743	1	< 0.1%
3471522	1	< 0.1%
3437283	1	< 0.1%
3370799	1	< 0.1%

open_acc_6m

Unsupported

MISSING

REJECTED
UNSUPPORTED

Missing

466285

Missing (%)

100.0%

Memory size

3.6 MiB

open_il_6m

Unsupported

MISSING

REJECTED
UNSUPPORTED

Missing

466285

Missing (%)

100.0%

Memory size

3.6 MiB

open_il_12m

Unsupported

MISSING

REJECTED
UNSUPPORTED

Missing

466285

Missing (%)

100.0%

Memory size

3.6 MiB

open_il_24m

Unsupported

MISSING

REJECTED
UNSUPPORTED

Missing

466285

Missing (%)

100.0%

Memory size

3.6 MiB

mths_since_rent_il

Unsupported

MISSING

REJECTED
UNSUPPORTED

Missing

466285

Missing (%)

100.0%

Memory size

3.6 MiB

total_bal_it		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

il_util		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

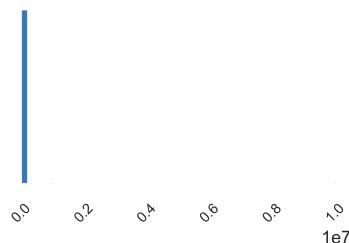
open_rv_12m		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

open_rv_24m		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

max_bal_be		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

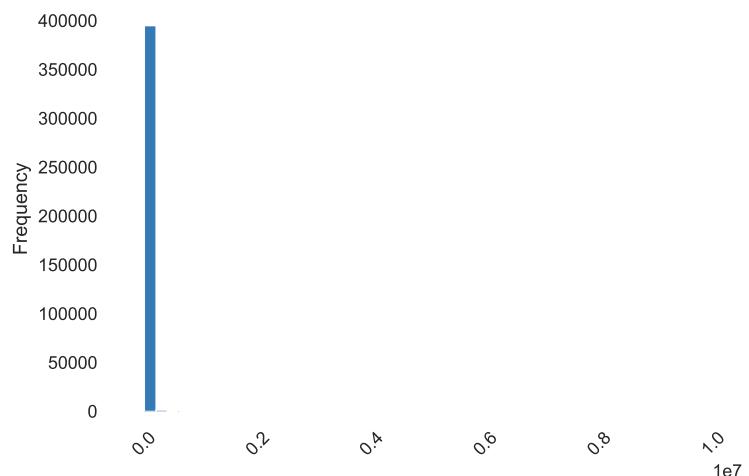
<u>all_util</u>	
<u>Unsupported</u>	MISSING
REJECTED	
UNSUPPORTED	
Missing	466285
Missing (%)	100.0%
Memory size	3.6 MiB

total_rev_hi_lim	
Real number ($\mathbb{R}_{\geq 0}$)	HIGH
CORRELATION (This variable has a high correlation with 1 fields: revol_bal)	
HIGH CORRELATION (This variable has a high correlation with 1 fields: revol_bal)	
HIGH CORRELATION (This variable has a high correlation with 1 fields: revol_bal)	
HIGH CORRELATION (This variable has a high correlation with 1 fields: revol_bal)	
MISSING	
SKEWED	
Distinct	14612
Distinct (%)	3.7%
Missing	70276
Missing (%)	15.1%
Infinite	0
Infinite (%)	0.0%
Mean	30379.08777
Minimum	0
Maximum	9999999
Zeros	155
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	3.6 MiB



Quantile statistics	
Minimum	0
5-th percentile	6000
Q1	13500
median	22800
Q3	37900
95-th percentile	77000
Maximum	9999999

Range	9999999
Interquartile range (IQR)	24400
Descriptive statistics	
Standard deviation	37247.12957
Coefficient of variation (CV)	1.226077947
Kurtosis	26001.4926
Mean	30379.08777
Median Absolute Deviation (MAD)	11000
Skewness	100.5855901
Sum	$1.203039217 \times 10^{10}$
Variance	1387348661
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
15000	1392	0.3%
13500	1324	0.3%
10000	1316	0.3%
12000	1301	0.3%
12500	1294	0.3%
14500	1290	0.3%
13000	1258	0.3%
11000	1257	0.3%
9000	1255	0.3%
11500	1254	0.3%
Other values (14602)	383068	82.2%
(Missing)	70276	15.1%

Value	Count	Frequency (%)
0	155	< 0.1%
100	7	< 0.1%
200	4	< 0.1%
300	30	< 0.1%
400	16	< 0.1%
500	84	< 0.1%
600	26	< 0.1%
605	1	< 0.1%
629	1	< 0.1%
696	1	< 0.1%

Value	Count	Frequency (%)
9999999	2	< 0.1%
2013133	1	< 0.1%
1998700	1	< 0.1%
1508600	1	< 0.1%
1314900	1	< 0.1%
1127900	1	< 0.1%
1090700	1	< 0.1%
1060800	1	< 0.1%
1030500	1	< 0.1%
1008700	1	< 0.1%

inq_hi		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	
total_cu_tt		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	
inq_last_12m		
Unsupported		MISSING
REJECTED		
UNSUPPORTED		
Missing	466285	
Missing (%)	100.0%	
Memory size	3.6 MiB	

Interactions

Unnamed: 0
id
member_id
loan_amnt
funded_amnt
funded_amnt_inv
int_rate
installment
annual_inc
dti
delinq_2yrs
inq_last_6mths
mths_since_last_delinq
mths_since_last_record
open_acc
pub_rec
revol_bal
revol_util
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
total_rec_int
total_rec_late_fee
recoveries
collection_recovery_fee
last_pymnt_amnt

total_rev_hi_lim
Unnamed: 0
id
member_id
loan_amnt
funded_amnt
funded_amnt_inv
int_rate
installment
annual_inc
dti
delinq_2yrs
inq_last_6mths
mths_since_last_delinq
mths_since_last_record
open_acc
pub_rec
revol_bal
revol_util
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
total_rec_int
total_rec_late_fee
recoveries
collection_recovery_fee

