

Data Summary

Name: ppp_orig
 Number of rows: 169268
 Number of columns: 24

Column type frequency:
 character: 17
 Date: 3
 numeric: 4

Group variables: None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
1 draw	0	1	5	6	0	2	0
2 borrower_name	0	1	1	60	0	144325	0
3 borrower_address	0	1	1	95	0	132559	0
4 borrower_city	0	1	2	32	0	562	0
5 borrower_state	0	1	2	2	0	7	0
6 borrower_fullzip	1	1.00	5	10	0	72792	0
7 borrower_zip	1	1.00	5	5	0	507	0
8 franchise_name	166541	0.0161	3	120	0	628	0
9 loan_status	0	1	11	19	0	3	0
10 lender	0	1	5	65	0	1113	0
11 rural_urban	0	1	1	1	0	2	0
12 low_income_area	0	1	1	1	0	2	0
13 project_county	20	1.00	4	10	0	20	0
14 project_state	0	1	2	2	0	6	0
15 project_cong_dist	15	1.00	3	5	0	20	0
16 business_type	42	1.00	5	35	0	23	0
17 naics_code	2587	0.985	6	6	0	1042	0

Provides an overview of the character columns, including the longest, shortest, empty, and unique values

For story purposes: Note that there are many fewer addresses than names - suggests more than one at an address.

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median
1 date_approved	0	1	2020-04-03	2021-06-29	2020-08-08
2 loan_status_date	56378	0.667	2020-06-30	2021-12-28	2021-07-01
3 forgiveness_date	36281	0.786	2020-10-07	2021-12-30	2021-07-01
n_unique					
1	235				
2	341				
3	295				

An overview of date columns, including the number empty, and the first and last

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	
1 loan_id	0	1	5381314427.	2612105285.	
2 amount	0	1	73214.	259961.	
3 forgiveness_amount	36281	0.786	78022.	266579.	
4 employees	0	1	9.32	29.9	
p0	p25	p50	p75	p100 hist	
1 1000067703	3043482527.	5338772055	7685712230	9998908810	█
2 10	9522	20800	47027.	10000000	█
3 1.01	10064.	20910.	54780.	10156667.	█
4 1	1	2	7	500	█

An overview of the numeric values, with basic statistics such as mean, standard deviation, and quantiles. (P50 is the same thing as "median")