

Column	NA	Exempt	NA/Exempt Category	Freeform Text	Drop Rows	Transformation after Dropping	Data Type	Original dataset: 383577 rows, 99 columns
activity_year					all		Categorical	
lei					all		ID	
derived_msa-md					all		ID	
state_code					all		Categorical	
county_code	2664				data=NA	replace code with text	Categorical	
census_tract					all		ID	
conforming_loan_limit	3781				data=NA	replace code with text	Categorical	
derived_loan_product_type						replace non-alphanumeric chars	Categorical	
derived_dwelling_category						replace non-alphanumeric chars	Categorical	
derived_ethnicity		83038		246	data="Ethnicity Not Available"	replace non-alphanumeric chars	Categorical	
derived_race		88456		158	data="Race Not Available"	replace non-alphanumeric chars	Categorical	
derived_sex		56478			data="Sex Not Available"	replace non-alphanumeric chars	Categorical	
action_taken						replace code with text	Categorical	
purchaser_type		258909				replace code with text	Categorical	
preapproval						change column to "preapproval_requested" and data to True/False	Categorical	
loan_type						replace code with text	Categorical	
loan_purpose		873			data=5	replace code with text	Categorical	
lien_status						change column to "secured_by_a_first_lien" and data to True/False	Boolean	
reverse_mortgate		10369			data=1111	change data to True/False	Boolean	
open-end_line_of_credit		10052			data=1111	change data to True/False	Boolean	
business_or_commercial_purpose		10351			data=1111	change column to "primarily_for_a_business_or_commercial_purpose" and data to True/False	Boolean	
loan_amount							Numerical	
loan_to_value_ratio	97202	10375			data="Exempt"	Impute missing numerical values	Numerical	
interest_rate	133511	10375			data="Exempt"	Impute missing numerical values	Numerical	
rate_spread	192300	10460			data="Exempt"	Impute missing numerical values	Numerical	
hoepa_status			182819		all		Boolean	
total_loan_costs	215684	10391			all		Numerical	
total_points_and_fees	372271	10391			all		Numerical	
origination_charges	214488	10361			all		Numerical	
discount_points	292732	10389			all		Numerical	
lender_credits	317307	10389			all		Numerical	
loan_term	11348	10023			data="Exempt"	Impute missing numerical values	Numerical	

Column	NA	Exempt	NA/Exempt Category	Freeform Text	Drop Rows	Transformation after Dropping	Data Type	Original dataset: 383577 rows, 99 columns
prepayment_penalty_term	354578	10394			all		Numerical	
intro_rate_period	272037	10374			all		Numerical	
negative_amortization			10379		data=1111	change data to True/False	Boolean	
interest_only_payment			10379		data=1111	change data to True/False	Boolean	
balloon_payment			10379		data=1111	change data to True/False	Boolean	
other_nonamortizing_features			10379		data=1111	change data to True/False	Boolean	
property_value	65219	10375			data="Exempt"	Impute missing numerical values change column name to "site_built" and data to True/False	Numerical	
construction_method						replace code with text	Boolean	
occupancy_type							Categorical	
manufactured_home_secured_property_type			NA: 170047, Exempt: 10306		all		Categorical	
manufactured_home_land_property_interest			NA: 170009, Exempt: 4631		all		Categorical	
total_units						use `df['total_units'] = df['total_units'].astype(str)` to resolve schema issue, replace data ">149" with "over_149"	Categorical	
multifamily_affordable_units	370235	10381			all		Numerical	
income	44847				data=NA	change column name to "income_in_k"	Numerical	
debt_to_income_ratio	109691	10375				drop "%" in data, change ">60" to "over_60"	Categorical	
applicant_credit_score_type			9971		all	replace code with text	Categorical	
co-applicant_credit_score_type			9971		all	replace code with text	Categorical	
applicant_ethnicity-1 through 5		varies				Use applicant_is_[ethnicity] columns and True/False as data	Categorical	
co-applicant_ethnicity-1 through 5		varies				Use co-applicant_is_[ethnicity] columns and True/False as data	Categorical	
applicant_ethnicity_observed			47146			change data to True/False/NA	Categorical	
co-applicant_ethnicity_observed			29657			change data to True/False/NA/no_co-applicant	Categorical	
applicant_race-1 through 5		varies				Use applicant_is_[race] columns and True/False as data	Categorical	
co-applicant_race-1 through 5		varies				Use co-applicant_is_[race] columns and True/False as data	Categorical	
applicant_race_observed			47078			change data to True/False/NA	Categorical	
co-applicant_race_observed			29427			change data to True/False/NA/no_co-applicant	Categorical	
applicant_sex			35765			change data to male/female/not_provided/NA/both	Categorical	
co-applicant_sex			24814			change data to male/female/not_provided/NA/no_co-applicant/both	Categorical	
applicant_sex_observed			47001			change data to True/False/NA	Categorical	
co-applicant_sex_observed			29396			change data to True/False/NA/no_co-applicant	Categorical	

Column	NA	Exempt	NA/Exempt Category	Freeform Text	Drop Rows	Transformation after Dropping	Data Type	Original dataset: 383577 rows, 99 columns
applicant_age			38951		data=8888	change "<25" to "below_25", ">74" to above_74	Categorical	
applicant_age_above_62			38951		data=NA	change data to True/False	Boolean	
co-applicant_age			247448		data=8888	change "<25" to "below_25", ">74" to above_74	Categorical	
co-applicant_age_above_62			247448		data=NA	change data to True/False	Boolean	
submission_of_application			NA: 27542, Exempt: 10005		all		Categorical	
initially_payable_to_institution			NA: 36647, Exempt: 10005		all		Categorical	
aus-1 through 5			NA: 198268, Exempt: 10411		data=1111	Use aus_is_[aus] columns and True/False as data	Categorical	
denial_reason_1 through 4			Exempt: 7360		data=1111	Use denial_reason_is_[denial_reason] columns and True/False as data	Categorical	
tract_population							Numerical	
tract_minority_population_percent							Numerical	
ffiec_msa_md_median_family_income							Numerical	
tract_to_msa_income_percentage							Numerical	
tract_owner_occupied_units							Numerical	
tract_one_to_four_family_homes							Numerical	
tract_median_age_of_housing_units							Numerical	