

# Codebook for EDA\_data

Autogenerated data summary from dataMaid

2020-08-26 16:14:38

## Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	1155
Number of variables	36

## Codebook summary table

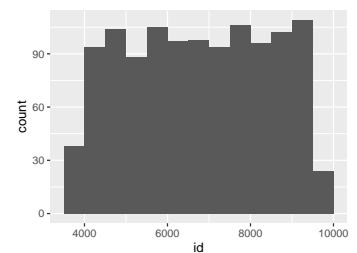
Label	Variable	Class	# unique values	Missing	Description
	id	numeric	1155	0.00 %	
	money_made_inv	numeric	1153	0.00 %	
	acc_now_delinq	factor	2	0.00 %	
	acc_open_past_24mths	numeric	23	0.00 %	
	addr_state	factor	45	0.00 %	
	annual_inc	numeric	286	0.00 %	
	application_type	factor	2	0.00 %	
	avg_cur_bal	numeric	1121	0.00 %	
	bc_util	numeric	679	0.00 %	
	delinq_2yrs	numeric	8	0.00 %	
	delinq_amnt	numeric	3	0.00 %	
	dti	numeric	974	0.00 %	
	earliest_cr_line	POSIXct	349	0.00 %	
	emp_length	factor	12	0.00 %	
	emp_title	factor	372	0.00 %	
	grade	factor	7	0.00 %	
	home_ownership	factor	4	0.00 %	
	initial_list_status	factor	2	0.00 %	
	int_rate	numeric	226	0.00 %	
	last_credit_pull_d	POSIXct	45	0.00 %	
	loan_amnt	numeric	199	0.00 %	
	mort_acc	numeric	11	0.00 %	
	num_sats	numeric	35	0.00 %	
	num_tl_120dpd_2m	factor	1	0.00 %	
	num_tl_90g_dpd_24m	numeric	6	0.00 %	
	num_tl_30dpd	factor	2	0.00 %	
	out_prncp_inv	numeric	501	0.00 %	
	pub_rec	numeric	7	0.00 %	
	pub_rec_bankruptcies	numeric	4	0.00 %	

Label	Variable	Class	# unique values	Missing	Description
	<b>purpose</b>	factor	12	0.00 %	
	<b>sub_grade</b>	ordered	33	0.00 %	
	<b>term</b>	factor	2	0.00 %	
	<b>tot_coll_amt</b>	numeric	156	0.00 %	
	<b>tot_cur_bal</b>	numeric	1155	0.00 %	
	<b>total_rec_late_fee</b>	numeric	32	0.00 %	
	<b>verification_status</b>	factor	3	0.00 %	

## Variable list

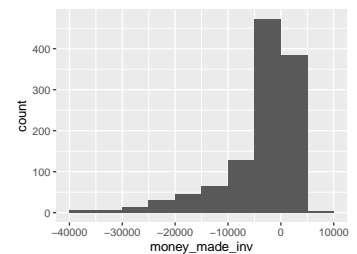
### id

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1155
Median	6779
1st and 3rd quartiles	5313.5; 8234.5
Min. and max.	3820; 9595



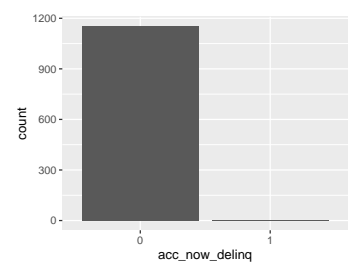
### money\_made\_inv

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1153
Median	-4.51
1st and 3rd quartiles	-5108.41; 1.96
Min. and max.	-38738.87; 9274.48



### acc\_now\_delinq

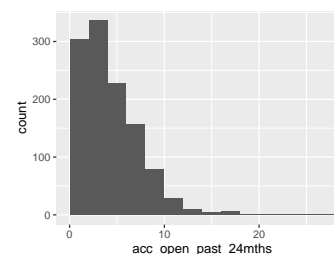
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



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

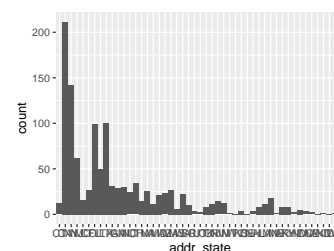
## acc\_open\_past\_24mths [Ordered => clip to 10+](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	23
Median	4
1st and 3rd quartiles	2; 6
Min. and max.	0; 28



## addr\_state [OOV](#)

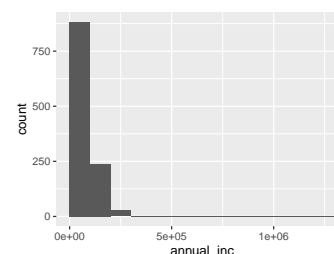
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	45
Mode	"CA"
Reference category	CT



- Observed factor levels: "AK", "AL", "AR", "AZ", "CA", "CO", "CT", "DC", "DE", "FL", "GA", "HI", "ID", "IL", "IN", "KS", "KY", "LA", "MA", "MD", "ME", "MI", "MN", "MO", "MS", "MT", "NC", "ND", "NE", "NH", "NJ", "NM", "NV", "NY", "OH", "OK", "OR", "PA", "RI", "SC", "TN", "TX", "UT", "VA", "VT", "WA", "WI", "WV", "WY".

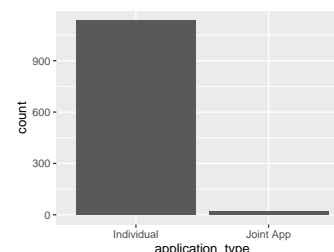
## annual\_inc [Log-scaling](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	286
Median	70000
1st and 3rd quartiles	50000; 1e+05
Min. and max.	10000; 1250007



## application\_type [stayed](#)

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

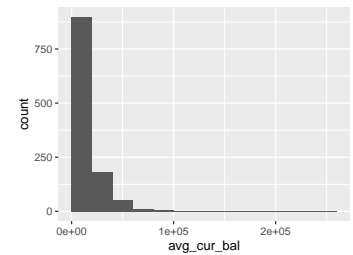


- Observed factor levels: "Individual", "Joint App".

---

## avg\_cur\_bal [Log-scaling](#)

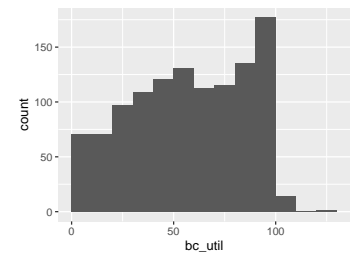
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1121
Median	7315
1st and 3rd quartiles	3197.5; 18711
Min. and max.	148; 251321



---

## bc\_util [Clipping to 100+ and Z-scoring](#)

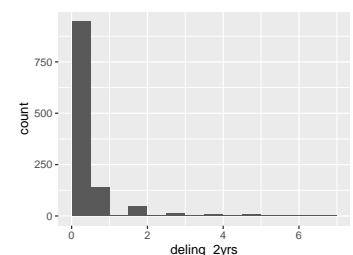
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	679
Median	57.8
1st and 3rd quartiles	34.3; 82.9
Min. and max.	0; 123.3



---

## delinq\_2yrs [to categorical -> OOV](#)

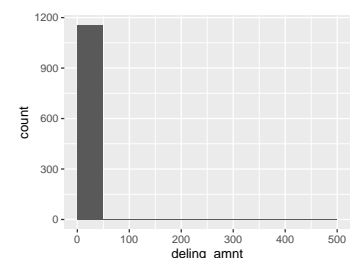
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	8
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 7



---

## delinq\_amnt [to categorical -> delinq or NOT\(0\)](#)

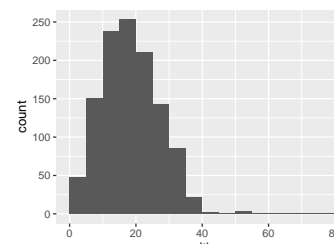
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	3
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 479



## dti

Clipping to 40+ and scaling to [0,1]

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	974
Median	17.65
1st and 3rd quartiles	12.11; 24.2
Min. and max.	0.16; 76.63



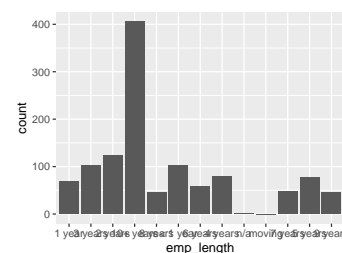
## earliest\_cr\_line

- The variable has class POSIXct which is not supported by dataMaid.

## emp\_length

<= 1 year, 2 years, 3-6 years, 7+ years, 10+ years

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	12
Mode	"10+ years"
Reference category	1 year

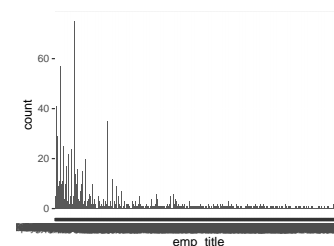


- Observed factor levels: "< 1 year", "1 year", "10+ years", "2 years", "3 years", "4 years", "5 years", "6 years", "7 years", "8 years", "9 years", "moving", "n/a".

## emp\_title

removed for now

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	372
Mode	"Teacher"
Reference category	Security Guard



- Observed factor levels: "Account Coordinator", "Account Director", "account executive", "Account executive", "Account Executive", "account manager", "Account manager", "Account Manager", "Account Representative", "accountant", "Accountant", "Accounting", "ACCOUNTING", "Accounting Assistant", "Accounting Clerk", "Accounting manager", "Accounting Manager", "Accounting Specialist", "Accounting Technician", "Accounts payable", "Accounts Payable", "accounts receivable", "Accounts Receivable", "Adjunct Professor", "Admin", "Admin Assistant", "Admin. Assist.", "Administration", "Administrative", "administrative assistant", "Administrative assistant", "Administrative Assistant", "Administrative Associate", "Administrative Officer", "Administrative Specialist", "administrator", "Administrator", "Advisor", "Agent", "Air Traffic Control Specialist", "Air Traffic Controller", "Aircraft Mechanic", "analyst", "Analyst", "Application Engineer", "Archaeologist", "Architect", "Area Manager", "Area Sales Manager", "ARNP", "Art Director", "Assembler", "Assembly", "Asset Protection Manager", "assistant", "Assistant",

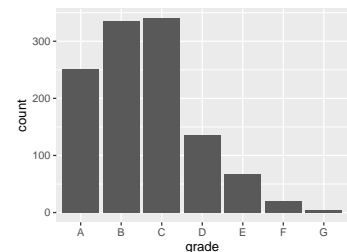
"Assistant Branch Manager", "Assistant Controller", "Assistant Director", "Assistant General Manager", "assistant manager", "Assistant manager", "Assistant Manager", "Assistant Manger", "Assistant Nurse Manager", "Assistant principal", "Assistant Principal", "Assistant Production Manager", "Assistant Professor", "Assistant Project Manager", "Assistant Store Manager", "Assistant Teacher", "Assistant Vice President", "Associate", "Associate Attorney", "Associate Director", "Associate Governmental Program Analyst", "Associate Producer", "Associate Professor", "AT&T", "Athletic Director", "attorney", "Attorney", "ATTORNEY", "Auditor", "auto body tech", "Auto Technician", "Avionics Technician", "AVP", "Baker", "Baker Hughes", "Bank Examiner", "Banker", "Banquet Captain", "bar manager", "bartender", "Bartender", "Behavior Intervention Specialist", "Billing", "BILLING MGR", "Billing Specialist", "Bookkeeper", "Border Patrol Agent", "branch manager", "Branch manager", "Branch Manager", "Broker", "Broker Associate", "Budget Analyst", "Building Maintenance", "building manager", "bus driver", "Bus driver", "Bus Driver", "bus operator", "Bus operator", "Bus Operator", "Business Analyst", "Business Consultant", "Business Development", "Business Development Director", "Business Development Manager", "Business Manager", "Business Operations Manager", "Business Systems Analyst", "Buyer", "Captain", "Car inspector", "Cardiac Monitor Tech", "caregiver", "Caregiver", "carpenter", "Carpenter", "case manager", "Case manager", "Case Manager", "Caseworker", "Cashier", "Category Manager", "CCO", "CEO", "Cerner Corporation", "CFO", "Chauffeur", "Checker", "Chef", "Chief Engineer", "Chief Financial Officer", "Chief Operating Officer", "Chief Technology Officer", "Child Protective Specialist", "CHMC Anesthesia Foundation", "CIO", "City Carrier", "Claims Adjuster", "Claims Analyst", "Claims Examiner", "Claims Supervisor", "Cleaner", "Clerical", "clerk", "Clerk", "Client Executive", "Client Manager", "Client Service Associate", "Client Service Manager", "Clinical Director", "Clinical Laboratory Scientist", "Clinical Nurse II", "Clinical Research Coordinator", "Clinical Social Worker", "Clinician", "CMO", "cna", "Cna", "CNA", "CNC Programmer", "Co-Owner", "Collector", "Commercial driver", "Communications Manager", "Communications Supervisor", "Compliance Manager", "Compliance Officer", "Compliance Specialist", "Comptroller", "Computing Infrastructure Engineer", "Concierge", "Conductor", "Construction Manager", "consultant", "Consultant", "Contract Administrator", "Contract Manager", "Contract Officer", "Contract Specialist", "Contractor", "Contracts Manager", "Controller", "COO", "cook", "Cook", "coordinator", "Coordinator", "Copywriter", "Corporate Accountant", "Correction officer", "Correction Officer", "correctional officer", "Correctional officer", "Correctional Officer", "Correctional Sergeant", "corrections officer", "Corrections officer", "Corrections Officer", "Counselor", "County of San Bernardino", "Courier", "CPA", "Crane operator", "Creative Director", "CRNA", "CSR", "CTO", "custodian", "Custodian", "Custody Officer", "Customer Advocate", "Customer Relationship Manager", "customer service", "Customer service", "Customer Service", "CUSTOMER SERVICE", "Customer Service Associate", "Customer Service Lead", "Customer Service Manager", "Customer service representative", "Customer Service Representative", "Customer Service Supervisor", "Customer Success Manager", "Data Analyst", "Database Administrator", "Dealer", "Delivery Driver", "Dental Assistant", "dental hygienist", "Dental hygienist", "Dental Hygienist", "Department Manager", "Department of Defense", "Deputy", "Deputy Clerk", "Deputy Director", "Deputy sheriff", "Deputy Sheriff", "Design Director", "Design Engineer", "Designer", "Detective", "Developer", "Dialysis Technician", "dining services director", "director", "Director", "DIRECTOR", "Director Change Leadership", "Director eCommerce", "Director of Business Development", "Director of Front Office", "Director of IT", "Director of Marketing", "Director of Nursing", "Director of Operations", "Director of Purchasing", "Director of Retail", "Director of Sales", "dispatcher", "Dispatcher", "District manager", "District Manager", "District Sales Manager", "District Secretary", "Doctor", "doorman", "Doorman", "driver", "Driver", "DRIVER", "Durability Test Driver", "E-5", "E-6", "Economist", "Editor", "Educator", "Electrical Engineer", "Electrical Foreman", "electrician", "Electrician", "Elementary Teacher", "Elevator Mechanic", "Employee Benefits Administrator", "EMT", "engineer", "Engineer", "Engineering Manager", "Engineering Tech", "Engineering Technician", "Enterprise Account Executive", "Environmental Specialist", "Equipment operator", "Escrow Assistant", "Estimator", "Event Specialist", "Evs", "Exec Assistant", "Executive", "Executive assistant", "Executive Assistant", "EXECUTIVE ASSISTANT", "Executive chef", "Executive Chef", "Executive Director", "Executive Recruiter", "Expediter", "fabricator", "Fabricator", "Facilities Manager", "Faculty", "Federal Agent", "feeder driver", "Feeder Driver", "Field Manager", "Field Nurse", "Field Service Engineer", "Field Service Tech", "Field Service Technician", "field technician", "Field Technician", "Finance", "Finance Director", "finance manager", "Finance Manager", "Financial advisor", "Financial Advisor", "Financial Aid Counselor", "Financial analyst", "Financial Analyst", "Financial Counselor", "Fire Captain", "firefighter", "Firefighter", "firefighter/paramedic", "Firefighter/Paramedic", "flight attendant", "Flight attendant", "Flight Attendant", "Flight Nurse", "foreman", "Foreman", "Forklift Driver", "forklift operator", "Forklift operator", "Forman", "Frac Operator", "Fraud Investigator", "Front desk", "Front Desk", "general manager", "General manager", "General Manager", "GENERAL MANAGER", "General Sales Manager", "Government Relations Specialist", "Grants Manager", "Graphic Designer", "Grocery Clerk", "Grocery manager", "Group Leader", "Guidance Counselor", "Hair stylist", "Haul Truck Driver", "Head teacher", "Head Teller", "heavy equipment operator", "Heavy Equipment Operator", "Heavy machine

operator", "Helen Newberry Joy Hospital", "HHA", "Host", "House Keeper", "House keeping", "Housekeeper", "Housekeeping", "Housing Specialist", "HR", "HR Director", "HR Manager", "HR Specialist", "Human Resource Manager", "Human Resources", "Human Resources Generalist", "Human Resources Manager", "HVAC Technician", "Implementation Consultant", "Infrastructure Analyst", "Inside Sales", "Inspector", "Installer", "Instructional Aide", "Instructional Assistant", "Instructional Specialist", "Instructional Support Technician", "Instructor", "insurance agent", "Insurance Agent", "Internal Auditor", "Investigator", "Investment Advisor", "IRS", "IT", "IT Administrator", "IT Analyst", "IT Consultant", "IT Director", "IT Engineer", "IT Manager", "IT Project Manager", "IT Specialist", "IT Tech", "kaiser permanente", "Kaiser Permanente", "KAISER PERMANENTE", "la county", "Lab assistant", "lab tech", "Lab tech", "Lab Technician", "labor", "Labor", "laborer", "Laborer", "LAUSD", "Law Enforcement", "Law Enforcement Officer", "Lead", "Lead Software Engineer", "Lead tech", "Legal assistant", "Legal Assistant", "Legal Secretary", "letter carrier", "Letter Carrier", "Librarian", "Licensed Massage Therapist", "Licensed practical nurse", "Licensed Practical Nurse", "Licensed vocational nurse", "Lieutenant", "Lineman", "Loan Officer", "Loan Processor", "Lobbyist", "Logistician", "Logistics", "Logistics Coordinator", "Logistics Manager", "Longshoreman", "Loudoun County Public Schools", "lpn", "LPN", "LTC", "LVN", "machine operator", "Machine operator", "Machine Operator", "Machinest", "machinist", "Machinist", "mail carrier", "Mail carrier", "Mail Carrier", "maintenance", "Maintenance", "Maintenance Manager", "Maintenance Mechanic", "Maintenance Specialist", "Maintenance Supervisor", "Maintenance technician", "Maintenance Technician", "Management", "Management Analyst", "manager", "Manager", "MANAGER", "Manager/owner", "Managing Director", "Managing Member", "managing partner", "Managing Partner", "manger", "Marketing", "Marketing Director", "Marketing Manager", "Mason", "Massage Therapist", "Material handler", "Material Handler", "Math Teacher", "MD", "meat cutter", "Meat Cutter", "mechanic", "Mechanic", "Mechanical Engineer", "medical assistant", "Medical assistant", "Medical Assistant", "Medical Assitant", "Medical Biller", "Medical Coder", "Medical Social Worker", "Medical Technologist", "Membership Manager", "Merchandiser", "Merchandising", "Mission Support Specialist", "Mortgage Loan Originator", "Mortgage Underwriter", "Motionographer", "Motor Vehicle Operator", "MRI Technologist", "MSA", "mvr", "Nanny", "Network Administrator", "Network Engineer", "Network Technical Specialist", "Network Technician", "nurse", "Nurse", "Nurse Assistant", "Nurse Consultant", "Nurse practitioner", "Nurse Practitioner", "Nursing", "Nursing assistant", "Nursing Assistant", "Occupational Therapist", "Office Administrator", "Office Assistant", "office manager", "office Manager", "Office manager", "Office Manager", "OFFICE MANAGER", "Office Support Assistant", "Office Technician", "officer", "Officer", "operating engineer", "Operating Engineer", "Operation Manager", "Operations", "Operations Analyst", "Operations Assistant", "Operations Director", "operations manager", "Operations manager", "Operations Manager", "OPERATIONS MANAGER", "Operations Supervisor", "operator", "Operator", "Optician", "Oracle", "Organizer", "Owned operator/truck driver", "owner", "Owner", "OWNER", "Owner/Operator", "Owner/President", "Pacific Gas and Electric", "painter", "Painter", "Para Educator", "Parae-ducator", "Paralegal", "Paramedic", "Paramedic Supervisor", "Paraprofessional", "partner", "Partner", "Partner - Attorney", "Passenger service agent", "pastor", "Pastor", "Pastry Chef", "Patent Examiner", "Patient Care Coordinator", "patient care technician", "Payroll Specialist", "PCA", "Peace Officer", "PepsiCo", "Perfusionist", "Personal Assistant", "Personal banker", "Personal Banker", "Personal Trainer", "pharmacist", "Pharmacist", "Pharmacy Manager", "Pharmacy tech", "Pharmacy Tech", "pharmacy technician", "Pharmacy technician", "Pharmacy Technician", "Phlebotomist", "Photographer", "Physical Therapist", "physician", "Physician", "Physician assistant", "Physician Assistant", "Physicians' Assistant", "Pilot", "Planner", "Plant Manager", "plumber", "Plumber", "Police", "police officer", "Police officer", "Police Officer", "Policy Analyst", "porter", "Porter", "Portfolio Manager", "Portland State University", "Postdoctoral Fellow", "Postmaster", "president", "President", "PRESIDENT", "President/CEO", "Press operator", "Press Operator", "Principal", "Principal Consultant", "Principal Engineer", "Process Engineer", "Process Operator", "Process Technician", "Procurement Manager", "Producer", "Product manager", "Product Manager", "production", "Production", "Production Control", "Production Coordinator", "Production Lead", "Production Manager", "Production Supervisor", "Production Worker", "Professor", "Program Analyst", "Program Coordinator", "Program Director", "Program Manager", "Program Specialist", "Programmer", "Project Coordinator", "Project Director", "Project Engineer", "Project Executive", "project manager", "Project manager", "Project Manager", "PROJECT MANAGER", "Project Specialist", "Property Management", "property manager", "Property Manager", "Psychologist", "PTA", "Public Safety", "Purchasing", "Purchasing Manager", "QA Manager", "QA Specialist", "Quality Analyst", "Quality Assurance", "Quality Control", "Quality Engineer", "Quality Supervisor", "Radiologic Technologist", "RDA", "Real estate agent", "Real Estate Agent", "Real Estate Sales", "Realtor", "Receiver", "Reception", "Receptionist", "Recruiter", "Regional Manager", "Regional Sale Consultant", "Regional Sales Manager", "Regional Supervisor", "Regional VP Sales", "Register nurse", "Registered Dental Hygienist", "registered nurse", "Registered nurse", "Registered Nurse", "REGISTERED NURSE", "Relationship Manager", "Research Assistant", "Research Associate", "Researcher", "Resident physician", "Resident Physician", "Respiratory therapist", "Respiratory

Therapist", "Retail Manager", "Retail Sales", "Risk Manager", "rn", "Rn", "RN", "RN Care Manager", "RN Case Manager", "rural carrier", "Rural Carrier", "Safety Manager", "sales", "Sales", "SALES", "Sales Administrator", "Sales agent", "sales associate", "Sales associate", "Sales Associate", "Sales Consultant", "Sales Coordinator", "Sales Director", "Sales Engineer", "Sales Executive", "sales manager", "Sales manager", "Sales Manager", "SALES MANAGER", "sales operations", "sales person", "sales rep", "Sales Representative", "Sales Specialist", "Salesman", "Sanitation worker", "School bus driver", "School Bus Driver", "School Counselor", "School Principal", "School Psychologist", "Scientist", "Secondary Teacher", "Secretary", "Security", "Security Engineer", "Security guard", "Security Guard", "Security Manager", "security officer", "Security officer", "Security Officer", "Senior Account Executive", "Senior Account Manager", "Senior Accountant", "Senior Administrative Assistant", "Senior Advisor", "Senior Analyst", "Senior Application Developer", "Senior Appraiser", "Senior Associate", "Senior Business Analyst", "Senior Buyer", "Senior Consultant", "Senior Director", "Senior Engineer", "Senior Estimator", "Senior Manager", "Senior Managing Consultant", "Senior Marketing Manager", "Senior Network Engineer", "Senior Pastor", "Senior Project Manager", "Senior Publicist", "Senior Relationship Banker", "Senior Sales Consultant", "Senior Sales Executive", "Senior Software Developer", "Senior Software Engineer", "Senior Technician", "Senior Vice President", "Senior Web Developer", "Sergeant", "Sergeant First Class", "Sergeant Major", "server", "Server", "Service Advisor", "service manager", "Service manager", "Service Manager", "service tech", "Service tech", "Service Tech", "service technician", "Service Technician", "Shareholder", "Shift Manager", "Shift supervisor", "Shift Supervisor", "Shipping", "Shop Foreman", "social worker", "Social worker", "Social Worker", "Software Consultant", "Software Developer", "Software Development Engineer", "software engineer", "Software engineer", "Software Engineer", "Solutions Manager", "Sous Chef", "Special Agent", "Special Education Teacher", "Specialist", "Speech Language Pathologist", "Speech Pathologist", "Sr Accountant", "Sr Consultant", "Sr Portfolio Administrator", "Sr Project Manager", "Sr. Analyst", "Sr. Financial Analyst", "Sr. Global Logistics Coordinator", "Sr. Manager", "Sr. Program Manager", "Sr. Project Manager", "Sr. Recruiter", "Sr. Systems Analyst", "Sr. Underwriter", "Staff Accountant", "Staff Attorney", "Staff Information Systems Analyst", "Staff nurse", "Staff Nurse", "Staff Pharmacist", "Staff Sergeant", "State of California", "State of NJ", "State Trooper", "station agent", "Stationary Engineer", "stocker", "Stocker", "store manager", "Store manager", "Store Manager", "stylist", "Stylist", "Substance Abuse Counselor", "super", "Superintendent", "Supervising RN", "supervisor", "Supervisor", "SUPERVISOR", "Surgical Technologist", "SVP", "System Administrator", "system analyst", "System Engineer", "Systems Admin", "Systems Administrator", "Systems Analyst", "Systems Engineer", "Table Games Dealer", "Talent Acquisition Specialist", "Target", "Tax Examiner", "Tax Manager", "teacher", "Teacher", "TEACHER", "Teacher Assistant", "Teacher/Coach", "Teachers", "Team lead", "Team Lead", "team leader", "Team Leader", "TEAM LEADER", "tech", "Tech", "Technical Account Manager", "technician", "Technician", "Teller", "Territory manager", "Therapist", "Toolmaker", "Tour Guide", "Train Dispatcher", "Train operator", "Train Operator", "Trainer", "Training Coordinator", "Training Specialist", "Transportation Security Specialist", "Trooper", "truck driver", "Truck driver", "Truck Driver", "Underwriter", "underwriting", "Union Carpenter", "Unit Supervisor", "United States Postal Service", "University of Illinois at Chicago", "University of Missouri", "UPMC", "US Army", "Utility Operator", "vice president", "Vice President", "VP", "VP of Operations", "VP of Sales", "VP Operations", "VP Sales", "vp-operations", "waiter", "Waiter", "Walmart", "warehouse", "Warehouse", "warehouse associate", "Warehouse Lead", "warehouse manager", "Warehouse manager", "Warehouse Manager", "Warehouse Specialist/Driver", "Water Foremen", "Web Design", "Web Developer", "Welder", "Wells Fargo Bank", "Winn Dixie", "Woodworker".

grade stayed

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



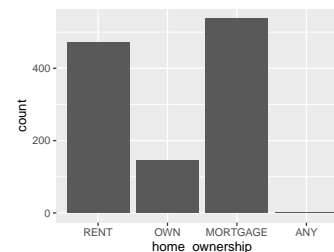
- Observed factor levels: "A", "B", "C", "D", "E", "F", "G".



## home\_ownership stayed

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

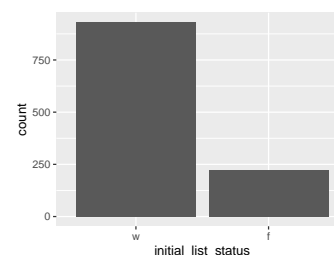
- Observed factor levels: "ANY", "MORTGAGE", "OWN", "RENT".



## initial\_list\_status stayed

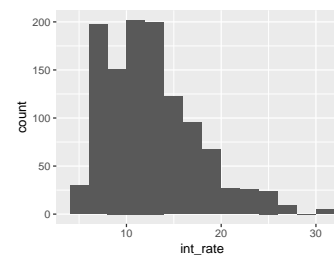
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"w"
Reference category	w

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



## int\_rate scaling to [0,1]

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	226
Median	11.99
1st and 3rd quartiles	9.17; 15.31
Min. and max.	5.31; 30.99

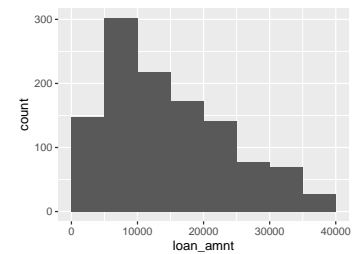


## last\_credit\_pull\_d

- The variable has class POSIXct which is not supported by dataMaid.

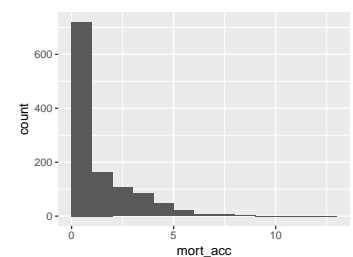
## loan\_amnt [Z-scoring](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	199
Median	13600
1st and 3rd quartiles	8000; 22000
Min. and max.	1000; 40000



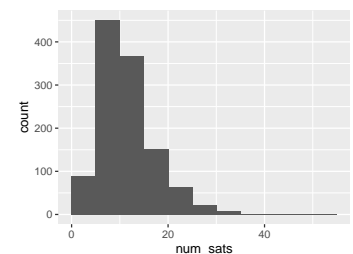
## mort\_acc [Clipping to max 6 => categorical](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	11
Median	1
1st and 3rd quartiles	0; 2
Min. and max.	0; 13



## num\_sats [Scaling to \[0,1\]](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	35
Median	11
1st and 3rd quartiles	8; 15
Min. and max.	2; 55

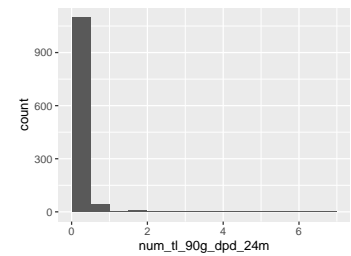


## num\_tl\_120dpd\_2m [OOV](#)

- The variable only takes one (non-missing) value: "0". The variable contains 0 % missing observations.

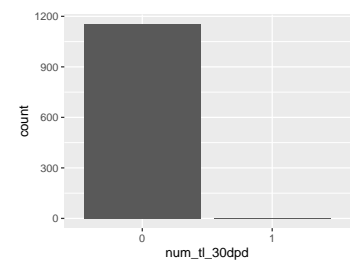
num\_tl\_90g\_dpd\_24m    oov

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	6
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 7



num\_tl\_30dpd    oov

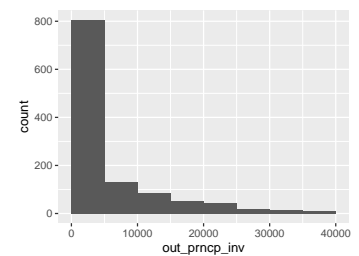
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



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

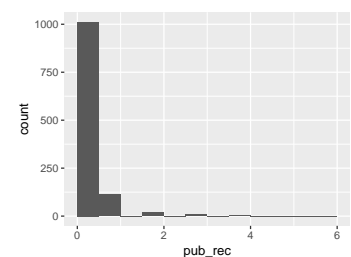
out\_prncp\_inv    Log-scaling

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	501
Median	0
1st and 3rd quartiles	0; 7271.6
Min. and max.	0; 38990.09



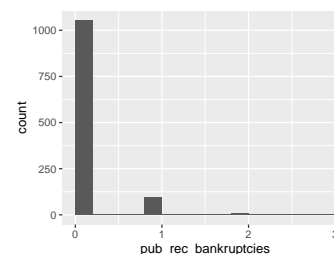
pub\_rec    oov  
categorical => 0, 1, 2+

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	7
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 6



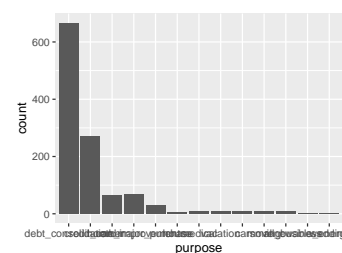
## pub\_rec\_bankruptcies   categorical => 0, 1, 2+

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



## purpose   OOV

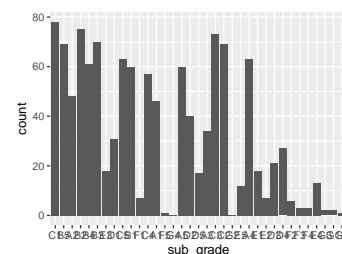
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	12
Mode	"debt_consolidation"
Reference category	debt_consolidation



- Observed factor levels: "car", "credit\_card", "debt\_consolidation", "home\_improvement", "house", "major\_purchase", "medical", "moving", "other", "renewable\_energy", "small\_business", "vacation", "wedding".

## sub\_grade   removed for now

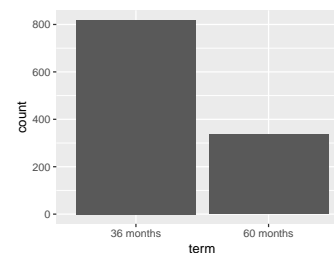
Feature	Result
Variable type	ordered
Number of missing obs.	0 (0 %)
Number of unique values	33
Mode	"C1"
Reference category	C1



- Observed factor levels: "A1", "A2", "A3", "A4", "A5", "B1", "B2", "B3", "B4", "B5", "C1", "C2", "C3", "C4", "C5", "D1", "D2", "D3", "D4", "D5", "E1", "E2", "E3", "E4", "E5", "F1", "F2", "F3", "F4", "F5", "G1", "G2", "G3", "G4", "G5".

## term   stayed

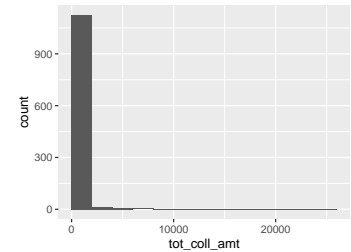
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"36 months"
Reference category	36 months



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

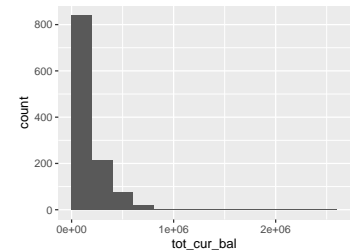
**tot\_coll\_amt** [try log-scaling and/or clipping, else turn to categorical](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	156
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 24597



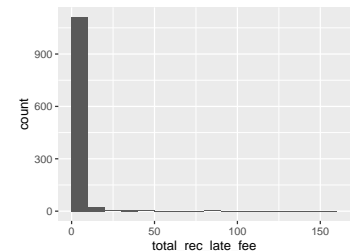
**tot\_cur\_bal** [Log-scaling](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1155
Median	83475
1st and 3rd quartiles	30458; 218797
Min. and max.	881; 2513207



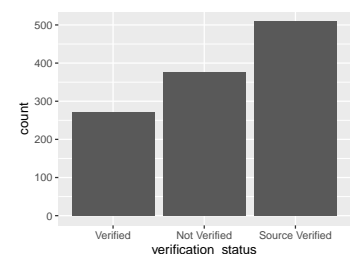
**total\_rec\_late\_fee** [turned to categorical => late or NOT\(0\)](#)

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	32
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 153



**verification\_status** [stayed](#)

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



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

---

Report generation information:

- Created by: Junhua Tan (username: JunhuaTan).
- Report creation time: Wed Aug 26 2020 16:14:39
- Report was run from directory: /Users/JunhuaTan/Documents/GitHub/STAT301-3-regression
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 3.6.0)]
- R version 3.6.3 (2020-02-29).
- Platform: x86\_64-apple-darwin15.6.0 (64-bit)(macOS Catalina 10.15.6).
- Function call: `dataMaid::makeDataReport(data = EDA_data, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook_EDA_data.Rmd", checks = list(character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for EDA_data")`