

EDA REPORT

Report Overview

This report was created for the EDA of . data. It helps explore data to **understand the data and find scenarios for performing the analysis.**

Contents

Overview	2
Data Structures	2
Job Informations	2
Univariate Analysis	3
Descriptive Statistics	3
Numerical Variables	3
Categorical Variables	5
Normality Test	9
Bivariate Analysis	12
Compare Numerical Variables	12
Compare Categorical Variables	14
Multivariate Analysis	15
Correlation Analysis	15
Correlation Coefficient Matrix	15
Correlation Plot	16

Overview

Data Structures

division	metrics	value	division	metrics	value
size	observations	12,487	data type	numerics	1
size	variables	14	data type	integers	1
size	values	174,818	data type	factors/ordered	0
size	memory size (MB)	2	data type	characters	12
duplicated	duplicate observation	0	data type	Dates	0
missing	complete observation	4,728	data type	POSIXcts	0
missing	missing observation	7,759	data type	others	0
missing	missing variables	9			
missing	missing values	69,712			

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	.
dataset	dataset type	data.frame
dataset	target	not defied
job	samples	12,487 / 12,487 (100%)
job	created	2022-02-17 10:17:28
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

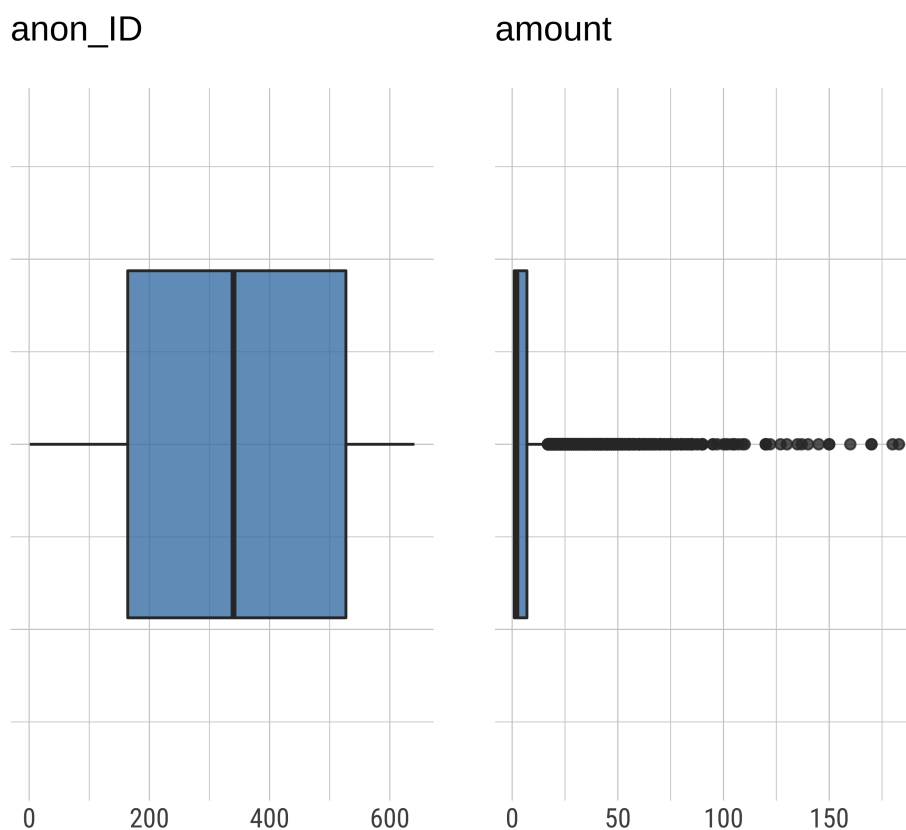
Descriptive Statistics

Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
anon_ID	0	334.15	197.93	1	164	340	527	641
amount	0	7.38	13.59	1	1	2	7	183

Table 3: Descriptive statistics of numerical variables

Distribution by numerical variables



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
anon_ID	integer	490	-0.07	-1.30	0	0	0
amount	numeric	113	4.19	25.82	0	0	1,481

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
assistance_date	2020-05-13 12:30:00	12,487	10	0.08	1
assistance_date	2020-07-17 16:00:00	12,487	10	0.08	1
assistance_date	2020-07-22 10:30:00	12,487	9	0.07	3
assistance_date	2020-04-07 16:00:00	12,487	7	0.06	4
assistance_date	2019-12-20 12:25:00	12,487	6	0.05	5
assistance_date	2020-04-27 11:30:00	12,487	6	0.05	5
assistance_date	2020-06-22 11:50:00	12,487	6	0.05	5
assistance_date	2020-08-11 15:00:00	12,487	6	0.05	5
assistance_date	2021-03-30 14:00:00	12,487	6	0.05	5
assistance_date	2019-12-09 15:30:00	12,487	5	0.04	10
assistance_category	Remote Contact with client (phone, email)	12,487	8,736	69.96	1
assistance_category	Remote Contact w/o client	12,487	1,948	15.60	2
assistance_category	Home Visit	12,487	766	6.13	3
assistance_category	Community	12,487	757	6.06	4
assistance_category	Internal Discussions	12,487	233	1.87	5
assistance_category	ElderNet Office	12,487	47	0.38	6
unit	Minutes	12,487	12,475	99.90	1
unit	Dollars	12,487	10	0.08	2
unit	Pounds	12,487	2	0.02	3
CommType	NA	12,487	7,742	62.00	1
CommType	Call	12,487	2,532	20.28	2
CommType	Voice Message	12,487	1,115	8.93	3
CommType	Email	12,487	694	5.56	4
CommType	In Person	12,487	152	1.22	5
CommType	Text Message	12,487	138	1.11	6

Table 4: Top rank levels of categorical variables

variables	levels	observations	frequency	frequency(%)	rank
variables	levels	observations	frequency	frequency(%)	rank
CommType		12,487	77	0.62	7
CommType	Mail	12,487	20	0.16	8
CommType	Video Call	12,487	15	0.12	9
CommType		12,487	2	0.02	10
Party	NA	12,487	7,742	62.00	1
Party	Client	12,487	3,272	26.20	2
Party	Service Provider	12,487	619	4.96	3
Party	Family	12,487	279	2.23	4
Party	ElderNet	12,487	232	1.86	5
Party	Other	12,487	126	1.01	6
Party	Social Worker	12,487	84	0.67	7
Party		12,487	70	0.56	8
Party	Friend	12,487	41	0.33	9
Party	Care Manager	12,487	12	0.10	10
InitiatedBy	NA	12,487	7,742	62.00	1
InitiatedBy	ElderNet	12,487	3,282	26.28	2
InitiatedBy	Other Party	12,487	1,100	8.81	3
InitiatedBy	Eldernet	12,487	203	1.63	4
InitiatedBy		12,487	100	0.80	5
InitiatedBy	Client	12,487	57	0.46	6
InitiatedBy	Service Provider	12,487	2	0.02	7
InitiatedBy	Other party	12,487	1	0.01	8
Benefit_1	NA	12,487	7,742	62.00	1
Benefit_1	ElderNet	12,487	1,547	12.39	2
Benefit_1		12,487	685	5.49	3
Benefit_1	Medical	12,487	638	5.11	4
Benefit_1	Housing	12,487	588	4.71	5
Benefit_1	Financial	12,487	407	3.26	6

Table 4: Top rank levels of categorical variables (continued)

variables	levels	observations	frequency	frequency(%)	rank
Benefit_1	Food	12,487	190	1.52	7
Benefit_1	Legal	12,487	146	1.17	8
Benefit_1	ADL	12,487	124	0.99	9
Benefit_1	Eldernet	12,487	118	0.94	10
Assistance_1	NA	12,487	7,742	62.00	1
Assistance_1	Coordination	12,487	2,443	19.56	2
Assistance_1		12,487	619	4.96	3
Assistance_1	Information	12,487	552	4.42	4
Assistance_1	Support	12,487	545	4.36	5
Assistance_1	Enrollment	12,487	289	2.31	6
Assistance_1	Continuation	12,487	117	0.94	7
Assistance_1	Filing	12,487	106	0.85	8
Assistance_1	Referral	12,487	71	0.57	9
Assistance_1	coordination	12,487	1	0.01	10
Benefit_2	NA	12,487	7,742	62.00	1
Benefit_2		12,487	4,154	33.27	2
Benefit_2	ElderNet	12,487	126	1.01	3
Benefit_2	Medical	12,487	108	0.86	4
Benefit_2	Financial	12,487	82	0.66	5
Benefit_2	Food	12,487	65	0.52	6
Benefit_2	Housing	12,487	64	0.51	7
Benefit_2	Legal	12,487	37	0.30	8
Benefit_2	Utilities	12,487	26	0.21	9
Benefit_2	ADL	12,487	23	0.18	10
Assistance_2	NA	12,487	7,742	62.00	1
Assistance_2		12,487	4,157	33.29	2
Assistance_2	Coordination	12,487	221	1.77	3
Assistance_2	Information	12,487	151	1.21	4

Table 4: Top rank levels of categorical variables (continued)

variables	levels	observations	frequency	frequency(%)	rank
Assistance_2	Support	12,487	118	0.94	5
Assistance_2	Enrollment	12,487	38	0.30	6
Assistance_2	Referral	12,487	30	0.24	7
Assistance_2	Filing	12,487	15	0.12	8
Assistance_2	Continuation	12,487	14	0.11	9
Assistance_2	ElderNet	12,487	1	0.01	10
Benefit_3	NA	12,487	7,759	62.14	1
Benefit_3		12,487	4,644	37.19	2
Benefit_3	ElderNet	12,487	20	0.16	3
Benefit_3	Financial	12,487	12	0.10	4
Benefit_3	Housing	12,487	10	0.08	5
Benefit_3	Legal	12,487	9	0.07	6
Benefit_3	ADL	12,487	5	0.04	7
Benefit_3	Food	12,487	5	0.04	7
Benefit_3	Medical	12,487	5	0.04	7
Benefit_3	Pets	12,487	5	0.04	7
Assistance_3	NA	12,487	7,759	62.14	1
Assistance_3		12,487	4,645	37.20	2
Assistance_3	Coordination	12,487	26	0.21	3
Assistance_3	Information	12,487	24	0.19	4
Assistance_3	Support	12,487	20	0.16	5
Assistance_3	Enrollment	12,487	5	0.04	6
Assistance_3	Referral	12,487	5	0.04	6
Assistance_3	Filing	12,487	2	0.02	8
Assistance_3	Continuation	12,487	1	0.01	9

Table 4: Top rank levels of categorical variables (continued)

The number of categorical(factor/ordered) variables is 0.

Normality Test

variable	min	Q1	median	Q3	max	skewness	kurtosis	balance
anon_ID	1	164	340	527	641	-0.1	-1.3	Balanced
amount	1	1	2	7	183	4.2	25.8	Right-Skewed

Table 5: Descriptive statistics of numerical variables

anon_ID

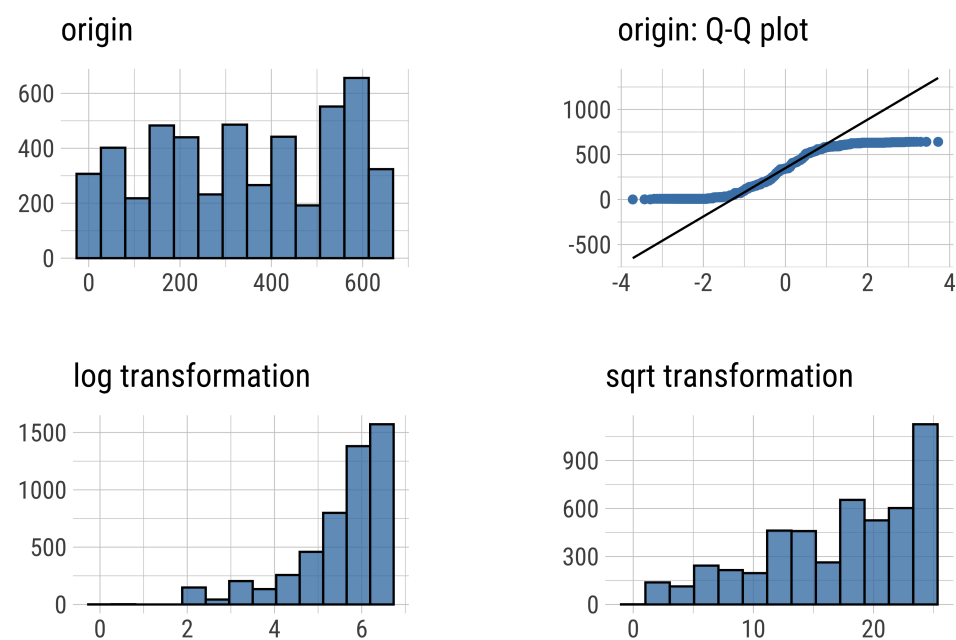
statistic	p_value	remark
0.93107	3.7178e-43	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.1072	1.6772
log transformation	-1.6080	5.2663
sqrt transformation	-0.6304	2.3362

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



amount

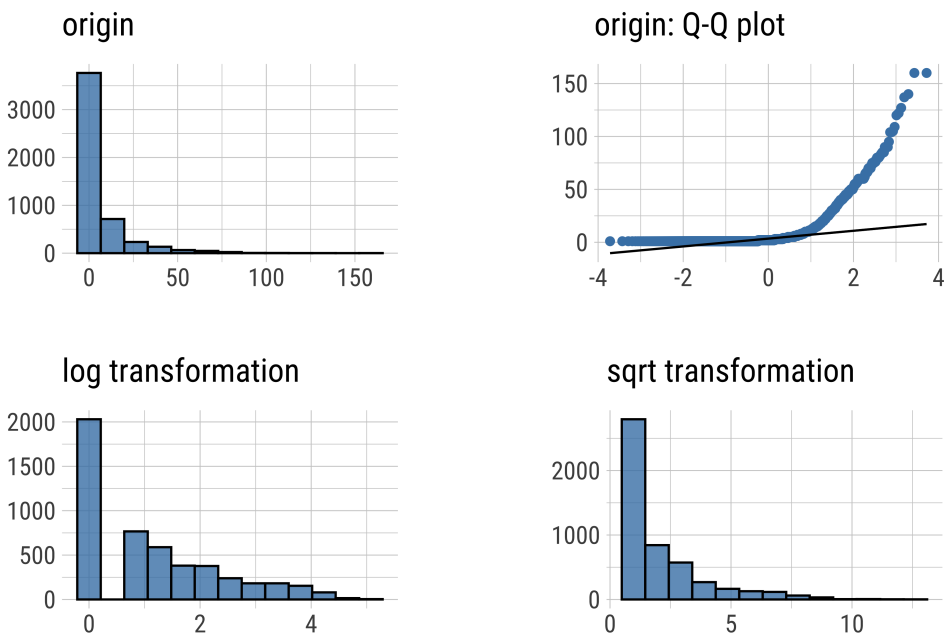
statistic	p_value	remark
0.50958	9.8831e-80	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	3.9390	24.3250
log transformation	0.9379	2.8676
sqrt transformation	2.1072	7.7136

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Bivariate Analysis

Compare Numerical Variables

first variable	second variable	correlation coefficient
anon_ID	amount	-0.04294

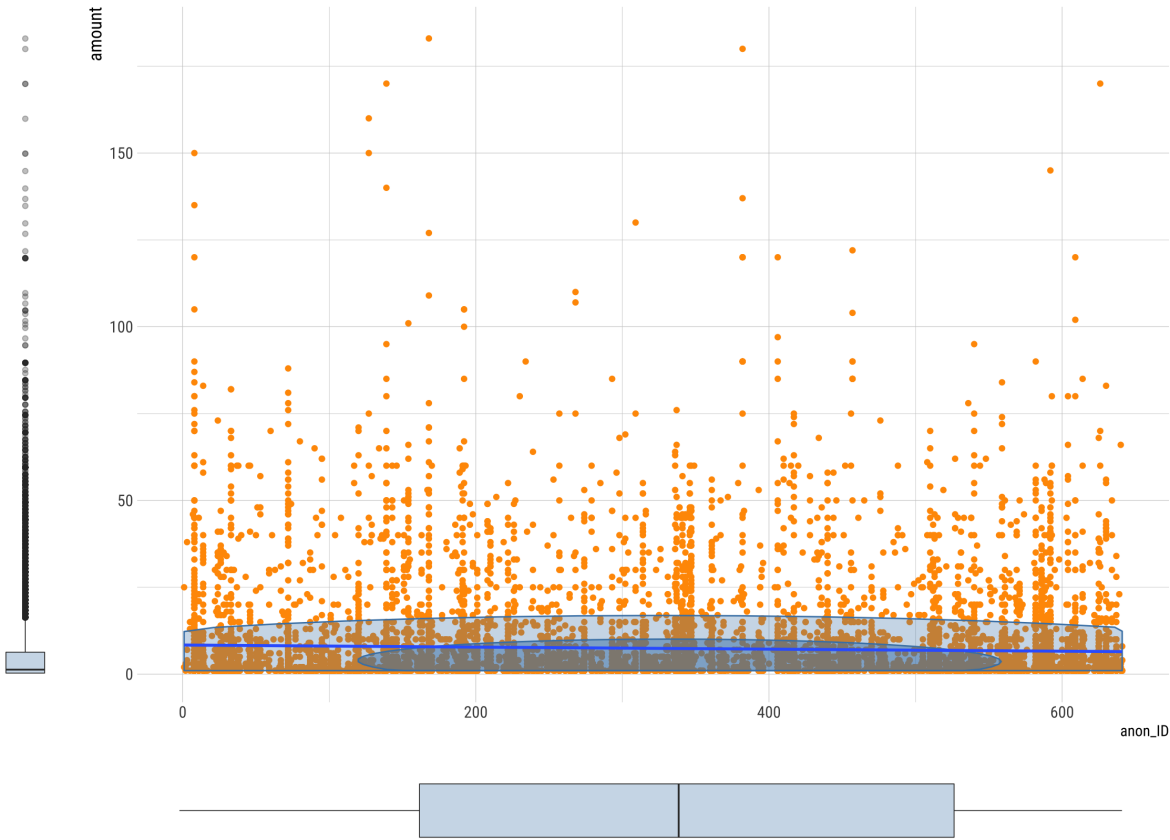
Table 7: Correlation coefficient

'anon_ID' vs 'amount'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
anon_ID	amount	0.0018435	0.0017635	197.7567	23.05819	0.0000016	1

Table 7: Summary of linear model

Scatterplots with anon_ID and amount



Compare Categorical Variables

The number of categorical variables is less than 2.

Multivariate Analysis

Correlation Analysis

Correlation Coefficient Matrix

first variable	second variable	
	anon_ID	amount
anon_ID	NA	-0.043
amount	-0.043	NA

Table 8: Matrix table of correlation coefficient

Correlation Plot

