EDA REPORT

Data by Nick Youngson CC BY-SA 3.0 Alpha Stock

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

Data Structures

division	metrics	value
size	observations	2,190
size	variables	11
size	values	24,090
size	memory size (KB)	0
duplicated	duplicate observation	2
missing	complete observation	2,054
missing	missing observation	136
missing	missing variables	3
missing	missing values	137

division	metrics	value
data type	numerics	2
data type	integers	0
data type	factors/ordered	0
data type	characters	7
data type	Dates	0
data type	POSIXcts	0
data type	others	2

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	.
dataset	dataset type	spec_tbl_df
dataset	target	not defied
job	samples	2,190 / 2,190 (100%)
job	created	2022-02-17 10:45:51
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

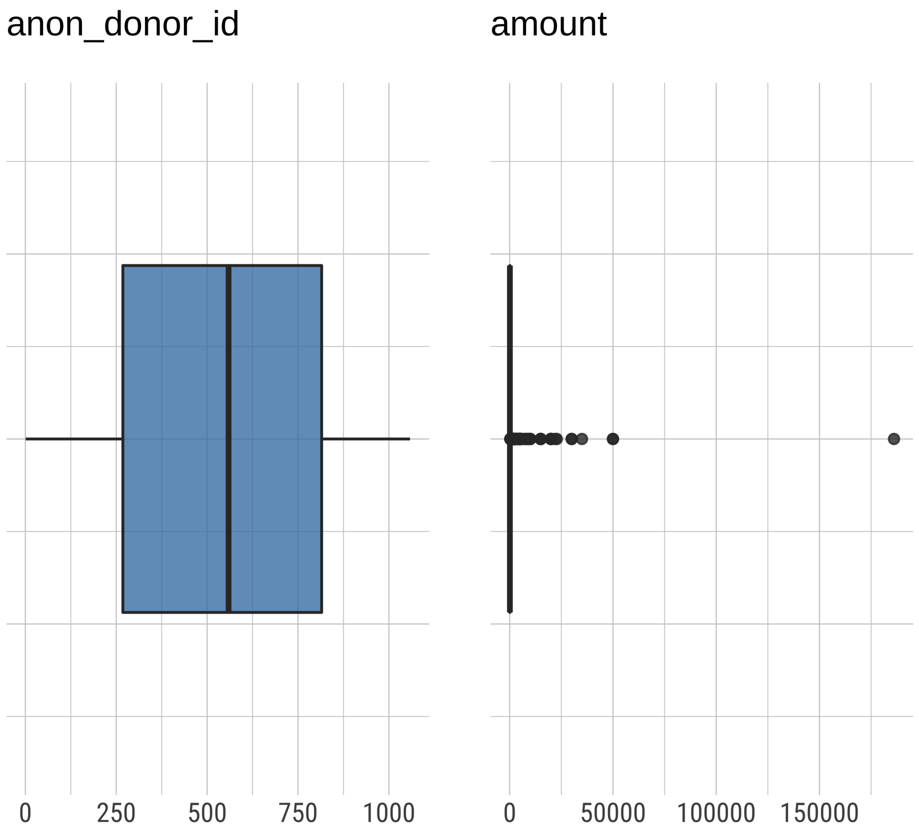
Descriptive Statistics

Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
anon_donor_id	0	541.32	308.34	1	268.25	559	815	1,058
amount	0	580.85	4,709.65	0	40.00	100	150	186,092

Table 3: Descriptive statistics of numerical variables

Distribution by numerical variables



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
anon_donor_id	numeric	1,058	-0.09	-1.24	0	0	0
amount	numeric	183	29.81	1,115.11	1	0	282

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
zip	19096	2,190	382	17.44	1
zip	19010	2,190	278	12.69	2
zip	19072	2,190	217	9.91	3
zip	19003	2,190	206	9.41	4
zip	19004	2,190	177	8.08	5
zip	19041	2,190	177	8.08	5
zip	19066	2,190	135	6.16	7
zip	NA	2,190	135	6.16	7
zip	19035	2,190	97	4.43	9
zip	19085	2,190	72	3.29	10
status	Active	2,190	2,189	99.95	1
status	NA	2,190	1	0.05	2
organisation	N	2,190	2,076	94.79	1
organisation	Y	2,190	114	5.21	2
date	12/11/2019	2,190	36	1.64	1
date	04/24/2020	2,190	29	1.32	2
date	05/05/2019	2,190	29	1.32	2
date	11/10/2020	2,190	25	1.14	4
date	07/28/2019	2,190	24	1.10	5
date	05/10/2021	2,190	23	1.05	6
date	10/16/2019	2,190	22	1.00	7
date	04/23/2020	2,190	19	0.87	8
date	05/01/2020	2,190	19	0.87	8
date	12/30/2019	2,190	19	0.87	8
form	Check	2,190	2,185	99.77	1

Table 4: Top rank levels of categorical variables

variables	levels	observations	frequency	frequency(%)	rank
variables	**levels**	**observations**	**frequency**	**frequency(%)**	**rank**
form	Cash	2,190	3	0.14	2
form	InKind	2,190	1	0.05	3
form	NA	2,190	1	0.05	3
campaign	Fall Towns	2,190	874	39.91	1
campaign	Sprg Evt	2,190	645	29.45	2
campaign	Misc	2,190	199	9.09	3
campaign	Emerg fund	2,190	130	5.94	4
campaign	Clients	2,190	69	3.15	5
campaign	Foundation	2,190	59	2.69	6
campaign	Board	2,190	41	1.87	7
campaign	Mem-hon	2,190	40	1.83	8
campaign	Church	2,190	29	1.32	9
campaign	Corporatio	2,190	22	1.00	10
target	Gift	2,190	2,190	100.00	1

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_donor_id	1	268.2	559	815	1058	-0.1	-1.2	Balanced
amount	0	40.0	100	150	186092	29.8	1115.1	Right-Skewed

Table 5: Descriptive statistics of numerical variables

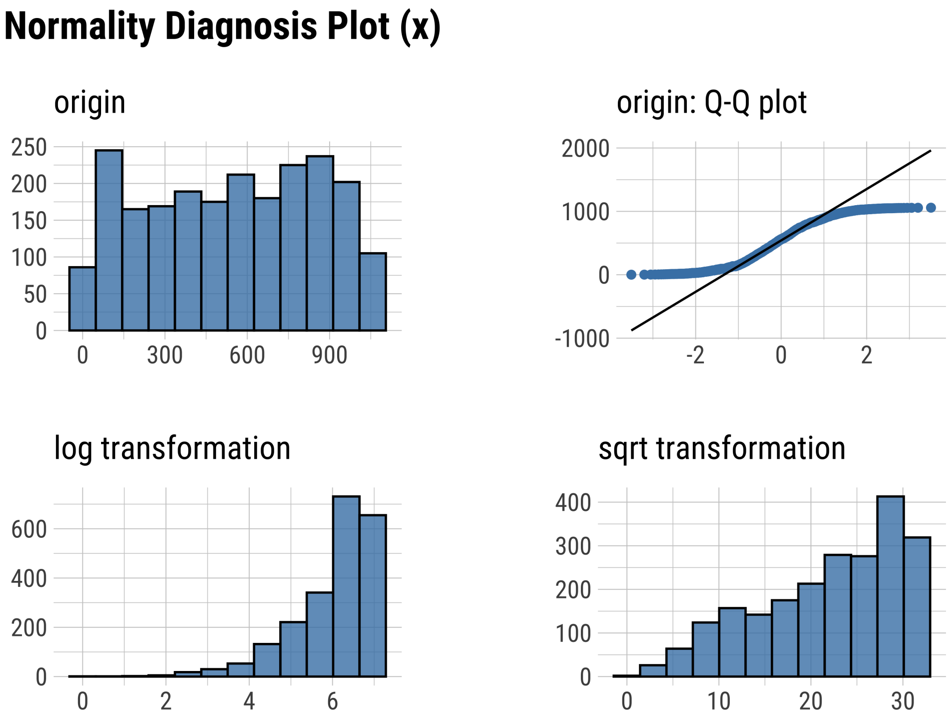
anon_donor_id

statistic	p_value	remark
0.9482	4.0757e-27	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.0861	1.7580
log transformation	-1.6416	6.2509
sqrt transformation	-0.5889	2.2933

Table 6: skewness and kurtosis



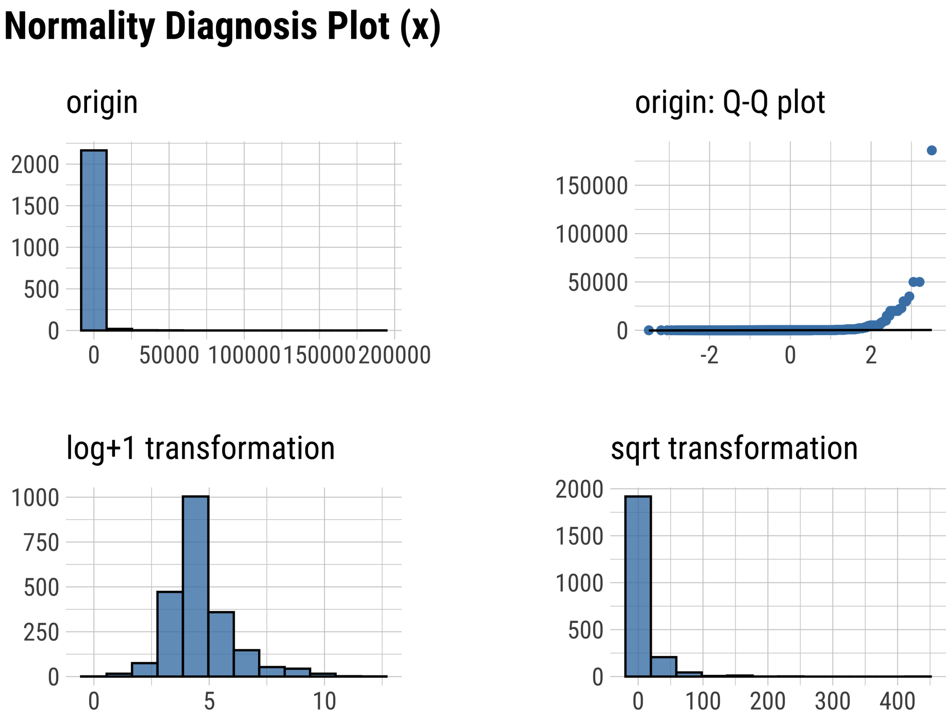
amount

statistic	p_value	remark
0.07607	2.5658e-73	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	29.7863	1115.5600
log+1 transformation	1.1905	5.7205
sqrt transformation	8.2279	117.3018

Table 6: skewness and kurtosis



Bivariate Analysis

Compare Numerical Variables

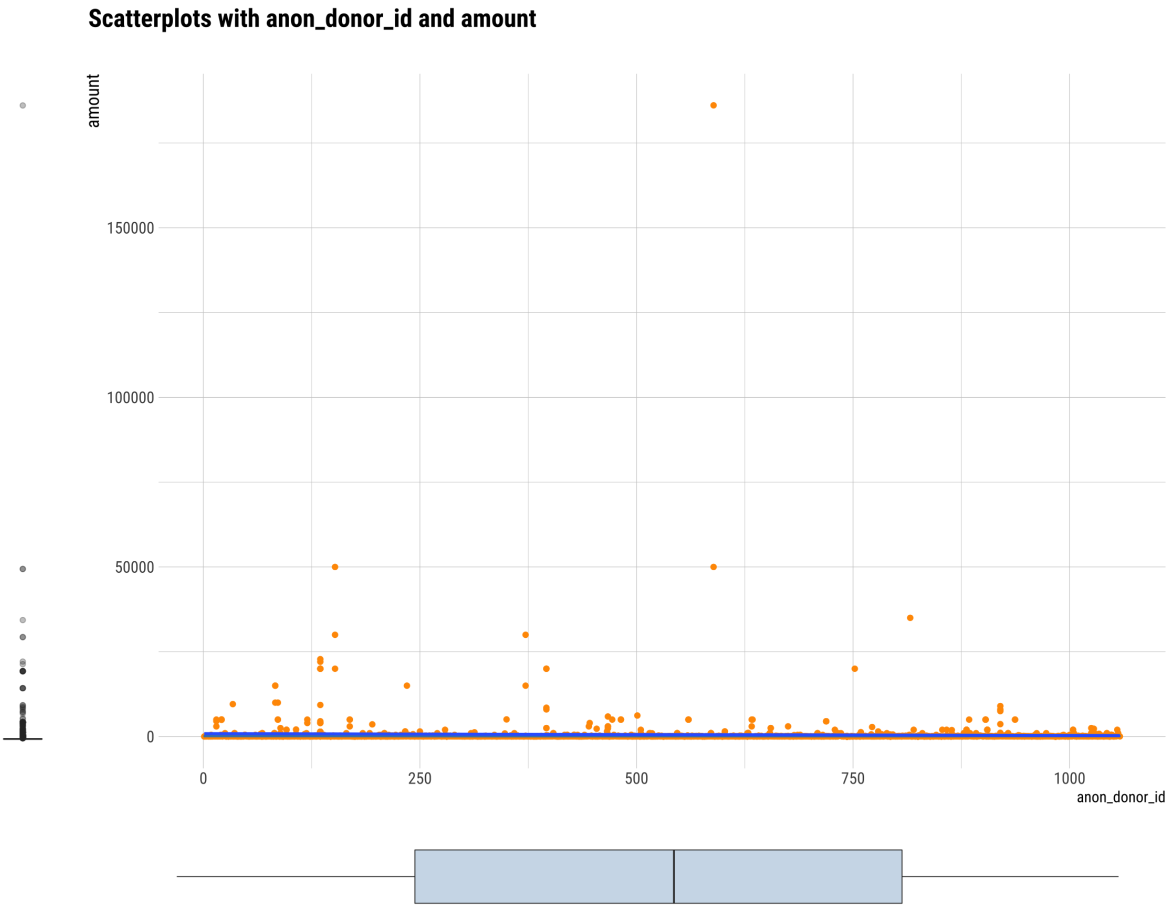
first variable	second variable	correlation coefficient
anon_donor_id	amount	-0.03476

Table 7: Correlation coefficient

'anon_donor_id' vs 'amount'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
anon_donor_id	amount	0.001208	0.0007515	308.2224	2.646355	0.1039316	1

Table 7: Summary of linear model



Compare Categorical Variables

The number of categorical variables is less than 2.

Multivariate Analysis

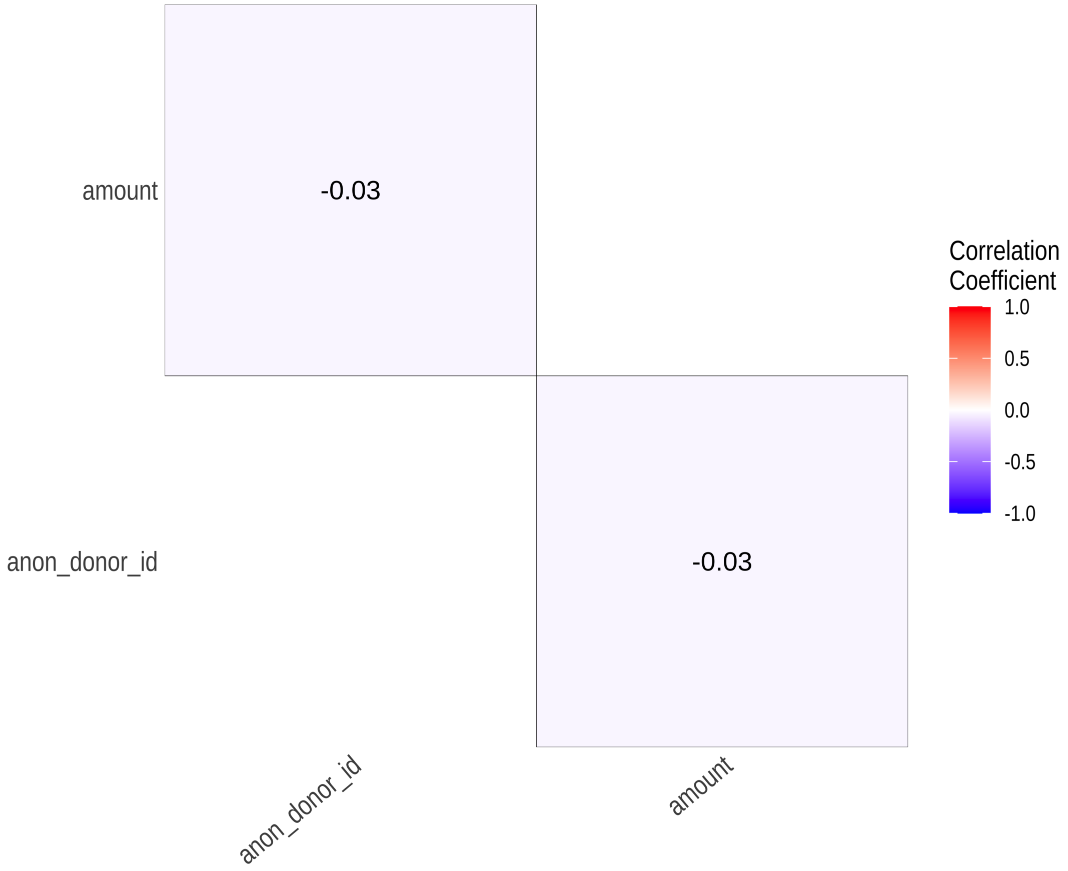
Correlation Analysis

Correlation Coefficient Matrix

	second variable	
first variable	anon_donor_id	amount
anon_donor_id	NA	-0.035
amount	-0.035	NA

Table 8: Matrix table of correlation coefficient

Correlation Plot


amount
anon_donor_id
anon_donor_id
amount
Correlation
Coefficient