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

Data Structures

division	metrics	value
size	observations	6,273
size	variables	5
size	values	31,365
size	memory size (MB)	1
duplicated	duplicate observation	132
missing	complete observation	6,273
missing	missing observation	0
missing	missing variables	0
missing	missing values	0

division	metrics	value
data type	numerics	2
data type	integers	0
data type	factors/ordered	0
data type	characters	3
data type	Dates	0
data type	POSIXcts	0
data type	others	0

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	6,273 / 6,273 (100%)
job	created	2022-02-17 10:56:28
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

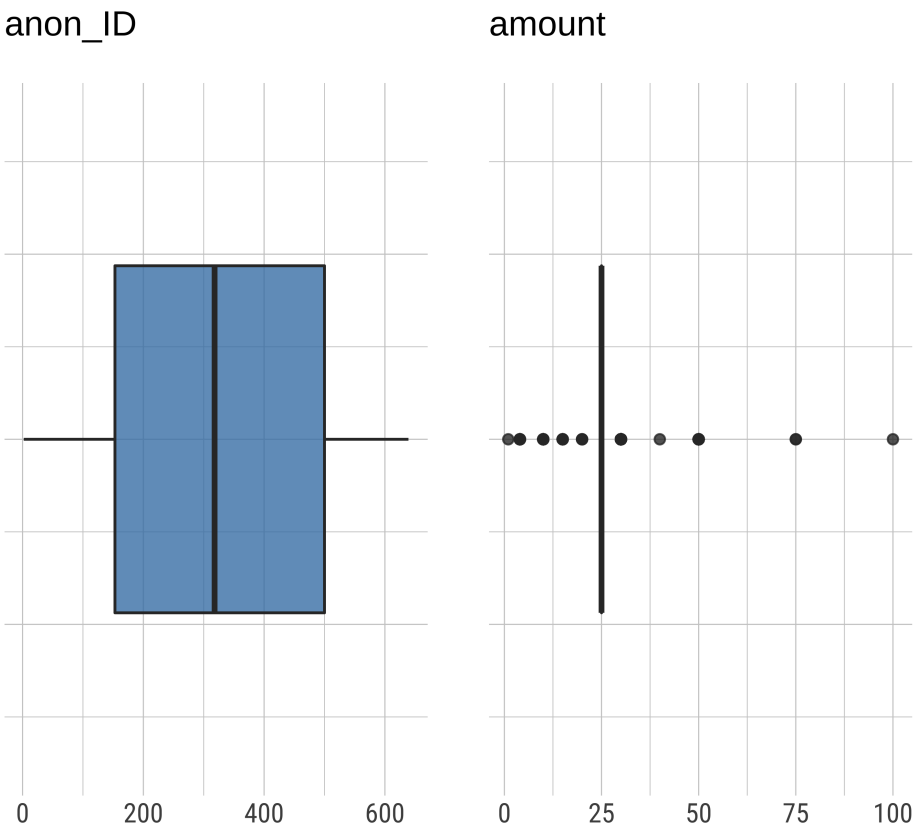
Descriptive Statistics

Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
anon_ID	0	319.1	191.73	2	153	318	500	639
amount	0	25.3	2.59	1	25	25	25	100

Table 3: Descriptive statistics of numerical variables

Distribution by numerical variables



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
anon_ID	numeric	414	-0.02	-1.24	0	0	0
amount	numeric	11	7.78	193.47	0	0	400

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
assistance_date	3/16/2020 13:35	6,273	13	0.21	1
assistance_date	3/16/2020 13:40	6,273	12	0.19	2
assistance_date	3/16/2020 13:20	6,273	11	0.18	3
assistance_date	5/26/2020 14:00	6,273	11	0.18	3
assistance_date	11/20/2020 12:00	6,273	10	0.16	5
assistance_date	12/22/2020 0:30	6,273	10	0.16	5
assistance_date	3/16/2020 13:25	6,273	10	0.16	5
assistance_date	10/27/2020 11:00	6,273	9	0.14	8
assistance_date	11/15/2019 16:00	6,273	8	0.13	9
assistance_date	12/22/2020 0:00	6,273	8	0.13	9
assistance_category	Food Pantry: Food Pantry Poundage	6,273	5,817	92.73	1
assistance_category	Food Pantry: Holiday Baskets	6,273	437	6.97	2
assistance_category	Food Pantry: Easter Outreach	6,273	19	0.30	3
unit	Pounds	6,273	6,161	98.21	1
unit	Dollars	6,273	99	1.58	2
unit	Boxes/Bags	6,273	13	0.21	3

Table 4: Top rank levels of categorical variables

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

Normality Test

variable	min	Q1	median	Q3	max	skewness	kurtosis	balance
anon_ID	2	153	318	500	639	0.0	-1.2	Balanced
amount	1	25	25	25	100	7.8	193.5	Right-Skewed

Table 5: Descriptive statistics of numerical variables

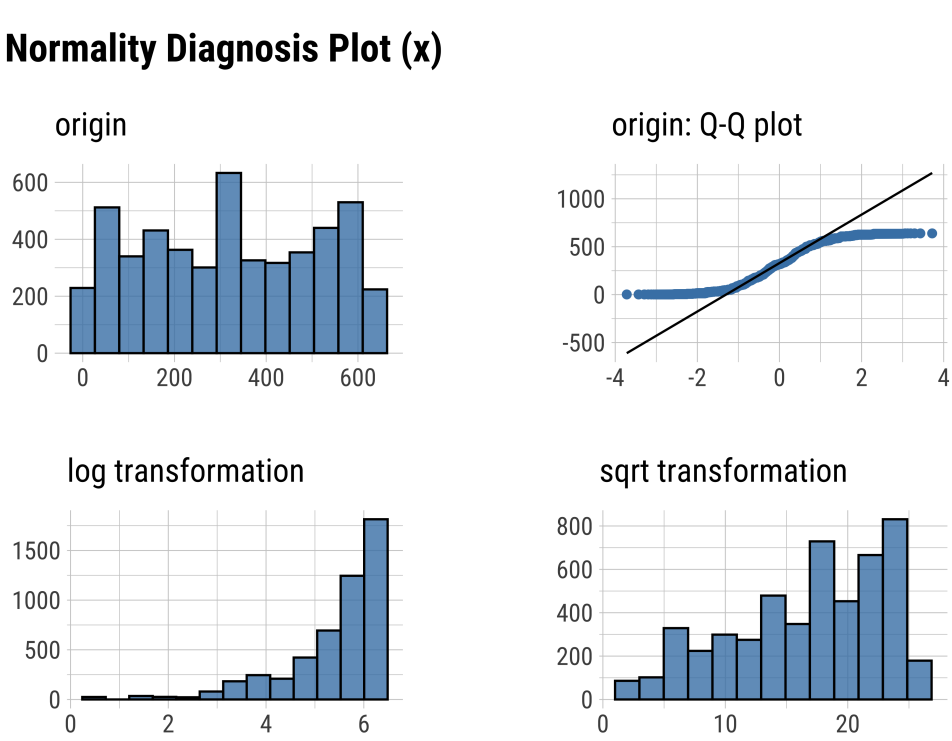
anon_ID

statistic	p_value	remark
0.94602	2.1097e-39	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.0171	1.7544
log transformation	-1.6633	5.9916
sqrt transformation	-0.5631	2.2879

Table 6: skewness and kurtosis



amount

statistic	p_value	remark
0.17989	3.2593e-91	5000 samples

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	10.5057	297.0612
log transformation	-8.2323	163.7251
sqrt transformation	0.6929	131.0159

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)

Bivariate Analysis

Compare Numerical Variables

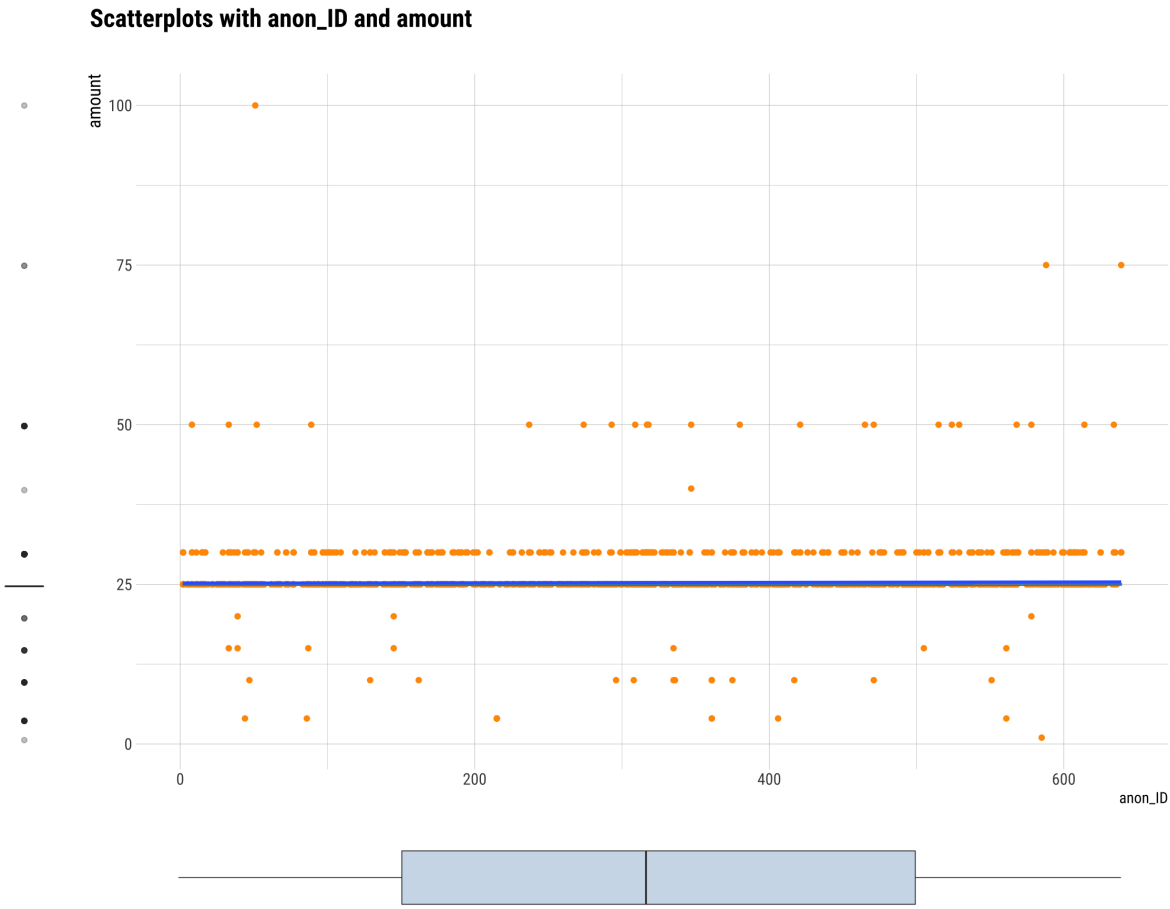
first variable	second variable	correlation coefficient
anon_ID	amount	0.0219

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	0.0004797	0.0003203	191.7038	3.009804	0.0828114	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_ID**	**amount**
anon_ID	NA	0.022
amount	0.022	NA

Table 8: Matrix table of correlation coefficient

Correlation Plot