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	948
size	variables	5
size	values	4,740
size	memory size (KB)	0
duplicated	duplicate observation	0
missing	complete observation	121
missing	missing observation	827
missing	missing variables	3
missing	missing values	1,251

division	metrics	value
data type	numerics	1
data type	integers	0
data type	factors/ordered	0
data type	characters	4
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	948 / 948 (100%)
job	created	2022-02-17 10:31:05
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

Descriptive Statistics

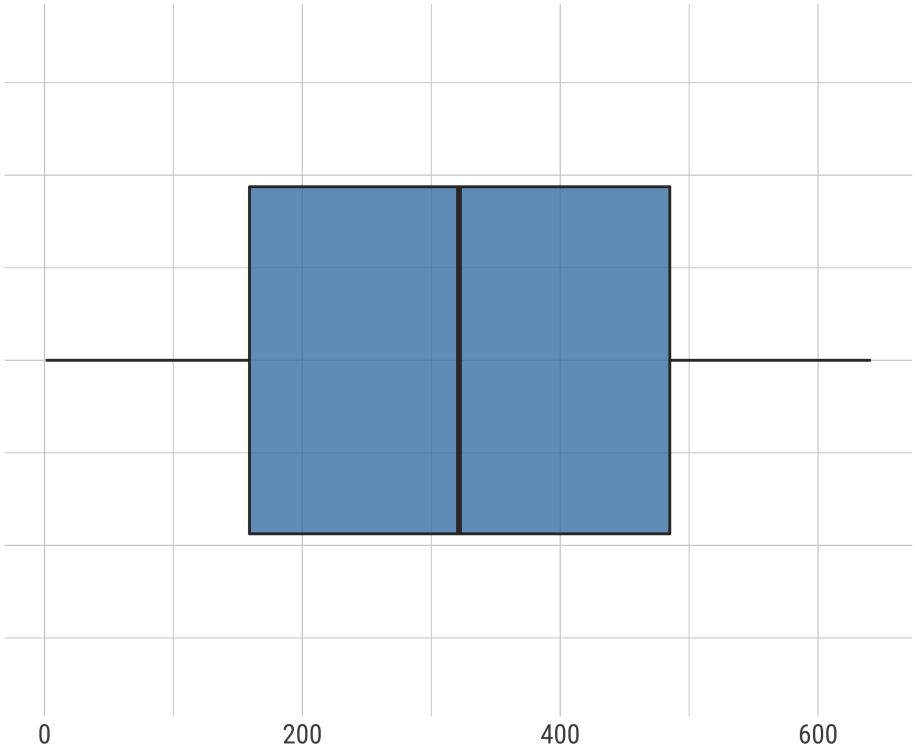
Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
anon_ID	0	321.48	187.39	1	159	321.5	485	641

Table 3: Descriptive statistics of numerical variables

Distribution by numerical variables

anon_ID



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
anon_ID	numeric	641	0	-1.22	0	0	0

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
county	Montgomery	948	908	95.78	1
county	Other	948	40	4.22	2
poverty	Yes	948	569	60.02	1
poverty	No	948	339	35.76	2
poverty	NA	948	40	4.22	3
minority	NA	948	713	75.21	1
minority	No	948	155	16.35	2
minority	Yes	948	80	8.44	3
age_group	NA	948	498	52.53	1
age_group	G	948	167	17.62	2
age_group	D	948	107	11.29	3
age_group	T	948	98	10.34	4
age_group	P	948	78	8.23	5

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	1	159	321.5	485	641	0	-1.2	Balanced

Table 5: Descriptive statistics of numerical variables

anon_ID

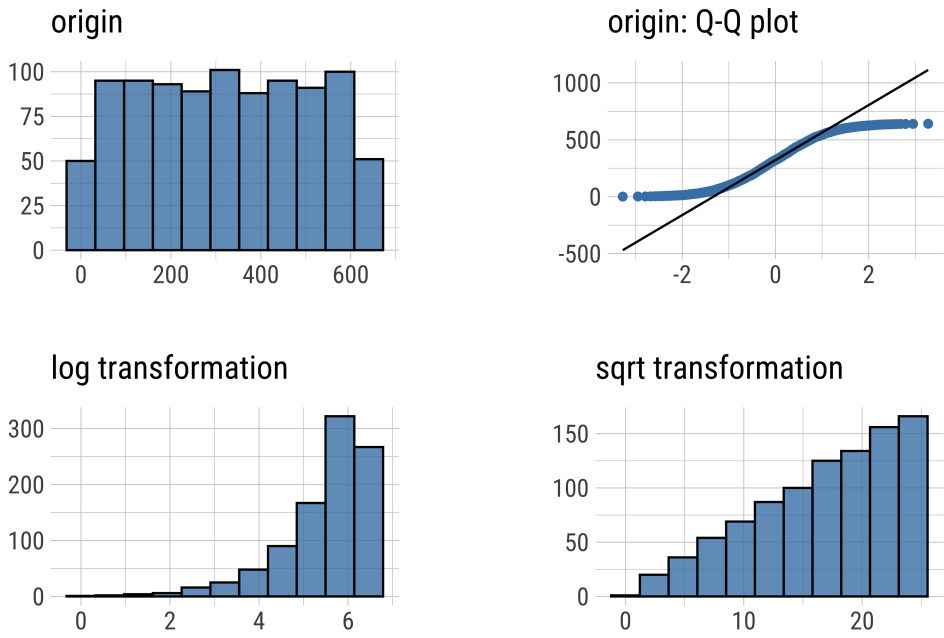
statistic	p_value	remark
0.95182	4.5504e-17	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.0022	1.7804
log transformation	-1.7392	6.6543
sqrt transformation	-0.5551	2.3465

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Bivariate Analysis

Compare Numerical Variables

The number of numerical variables is less than 2.

Compare Categorical Variables

The number of categorical variables is less than 2.

Multivariate Analysis

Correlation Analysis

Correlation Coefficient Matrix

The number of numerical variables is less than 2.

Correlation Plot

The number of numerical variables is less than 2.