EDA REPORT
SEPSIS

Report Overview

This report was created for the EDA of ***sepsis*** 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	456
size	variables	9
size	values	4,104
size	memory size (KB)	0
duplicated	duplicate observation	0
missing	complete observation	144
missing	missing observation	312
missing	missing variables	3
missing	missing values	344

division	metrics	value
data type	numerics	0
data type	integers	0
data type	factors/ordered	0
data type	characters	9
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	sepsis
dataset	dataset type	data.frame
dataset	target	not defied
job	samples	456 / 456 (100%)
job	created	2023-06-05 11:44:17.78944
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

Descriptive Statistics

Numerical Variables

There are no numerical variables.

The number of numerical variables is 0.

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
V1	1	456	1	0.22	1
V1	10	456	1	0.22	1
V1	100	456	1	0.22	1
V1	101	456	1	0.22	1
V1	102	456	1	0.22	1
V1	103	456	1	0.22	1
V1	104	456	1	0.22	1
V1	105	456	1	0.22	1
V1	106	456	1	0.22	1
V1	107	456	1	0.22	1
V2	Placebo	456	231	50.66	1
V2	Ibuprofen	456	224	49.12	2
V2	treat	456	1	0.22	3
V3	White	456	293	64.25	1
V3	African American	456	130	28.51	2
V3	Other	456	32	7.02	3
V3	race	456	1	0.22	4
V4	Alive	456	279	61.18	1
V4	Dead	456	176	38.60	2
V4	fate	456	1	0.22	3
V5	9	456	33	7.24	1
V5	13	456	32	7.02	2
V5	11	456	31	6.80	3
V5	17	456	27	5.92	4
V5	14	456	25	5.48	5

Table 3: Top rank levels of categorical variables

variables	levels	observations	frequency	frequency(%)	rank
variables	levels	observations	frequency	frequency(%)	rank
V5	16	456	24	5.26	6
V5	8	456	22	4.82	7
V5	10	456	19	4.17	8
V5	18	456	19	4.17	8
V5	15	456	18	3.95	10
V6	NA	456	287	62.94	1
V6	1002.119995	456	1	0.22	2
V6	1010.809998	456	1	0.22	2
V6	1014.299988	456	1	0.22	2
V6	1018.030029	456	1	0.22	2
V6	1019.669983	456	1	0.22	2
V6	1019.830017	456	1	0.22	2
V6	1028.170044	456	1	0.22	2
V6	1043.02002	456	1	0.22	2
V6	1045.578979	456	1	0.22	2
V7	720	456	279	61.18	1
V7	2	456	4	0.88	2
V7	22	456	3	0.66	3
V7	50	456	3	0.66	3
V7	51	456	3	0.66	3
V7	1	456	2	0.44	6
V7	10	456	2	0.44	6
V7	13	456	2	0.44	6
V7	165	456	2	0.44	6
V7	193	456	2	0.44	6
V8	38.00000085	456	19	4.17	1
V8	38.20000119	456	18	3.95	2
V8	38.5000017	456	18	3.95	2

Table 3: Top rank levels of categorical variables (continued)

variables	levels	observations	frequency	frequency(%)	rank
V8	38.10000102	456	13	2.85	4
V8	37.80000051	456	11	2.41	5
V8	38.40000153	456	11	2.41	5
V8	37.70000034	456	10	2.19	7
V8	37.77777778	456	10	2.19	7
V8	38.11111026	456	10	2.19	7
V8	38.30000136	456	10	2.19	7
V9	NA	456	56	12.28	1
V9	36.99999915	456	16	3.51	2
V9	37.5	456	16	3.51	2
V9	36.4999983	456	12	2.63	4
V9	37.61110942	456	11	2.41	5
V9	36.88888974	456	10	2.19	6
V9	36.89999898	456	10	2.19	6
V9	37.29999966	456	10	2.19	6
V9	37.60000017	456	10	2.19	6
V9	38.00000085	456	9	1.97	10

Table 3: Top rank levels of categorical variables (continued)

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

Normality Test

There are no numerical variables.

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.