



DATA DIAGNOSIS REPORT

SEPSIS_DF

Report Overview

This report was created for an overview quality diagnosis of *sepsis_df* data. It was created for the purpose of judging the validity of variables before conducting EDA.

Contents

Overview	2
Data Structures	2
Job Informations	2
Warnings	3
Variables	4
Missing Values	5
List of Missing Values	5
Visualization	6
Unique Values	7
Categorical Vaiables	7
Numerical Vaiables	8
Categorical Variable Diagnosis	9
Top Ranks	9
Numerical Variable Diagnosis	10
Distributions	10
Zero Values	11
Negative Values	12
Outliers	13
List of Outliers	13
Individual Outliers	14

Overview

Data Structures

division	metrics	value	division	metrics	value
size	observations	455	data type	numerics	6
size	variables	9	data type	integers	0
size	values	4,095	data type	factors/ordered	0
size	memory size (KB)	0	data type	characters	3
duplicated	duplicate observation	0	data type	Dates	0
missing	complete observation	143	data type	POSIXcts	0
missing	missing observation	312	data type	others	0
missing	missing variables	3			
missing	missing values	344			

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	sepsis_df
dataset	dataset type	spec_tbl_df
job	samples	455 / 455 (100%)
job	created	2023-11-27 17:14:39.019038
job	created by	dlookr

Table 2: Job informations

Warnings

checks	judgements	removes
2	7	0

Table 3: Summary of warnings

warnings	status	recommand
o2del has 287 (63.1%) missing values	missing	judgement
temp10 has 56 (12.3%) missing values	missing	judgement
apache has 1 (0.2%) missing values	missing	judgement
id has high(1.00) cardinality, Maybe identifier	cardinality	check
apache has 1 (0.22%) zeros	zero	check
temp0 has 28 (6.15%) outliers	outlier	judgement
o2del has 7 (1.54%) outliers	outlier	judgement
temp10 has 5 (1.1%) outliers	outlier	judgement
apache has 3 (0.66%) outliers	outlier	judgement

Table 4: Warnings in dataset and variables

Variables

variables	types	missing	cardinality	zero	minus	outlier
id	numeric		identifier			
treat	character					
race	character					
fate	character					
apache	numeric	X		X		X
o2del	numeric	X				X
followup	numeric					
temp0	numeric					X
temp10	numeric	X				X

Table 5: List of variables diagnosis

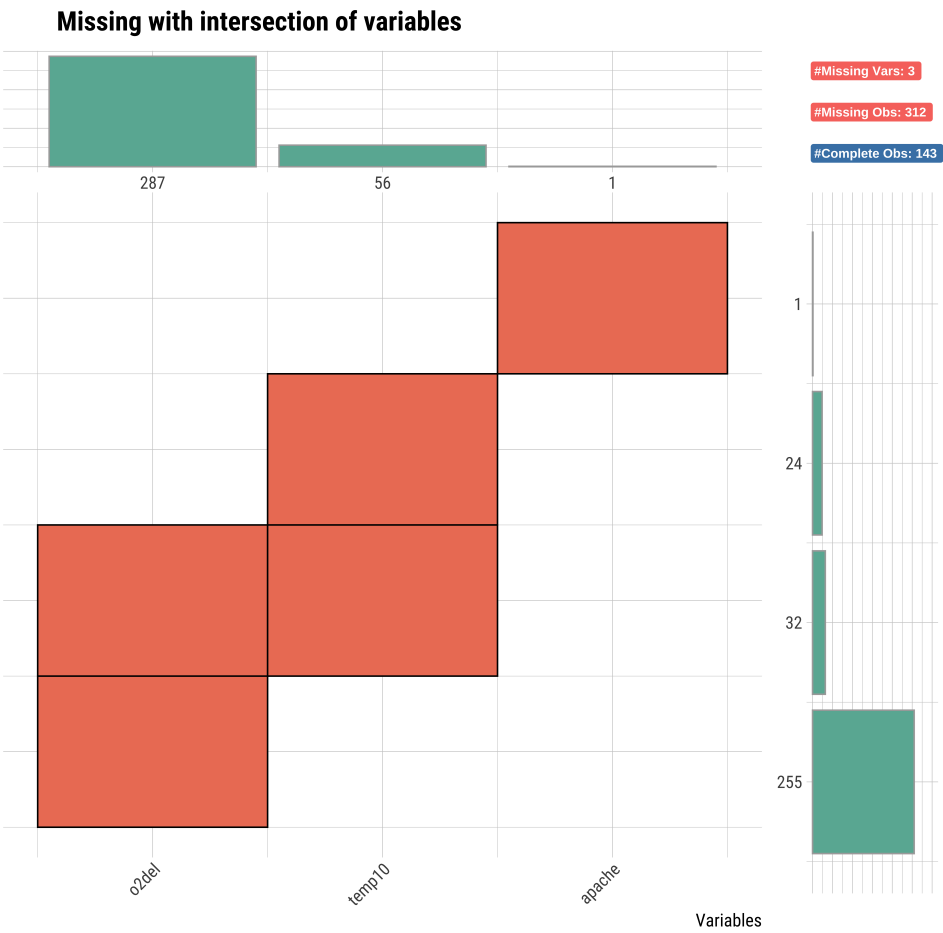
Missing Values

List of Missing Values

variables	missing_count	missing (%)	status	recommand
o2del	287	63.1%	Remove	Remove Variable
temp10	56	12.3%	NotBad	Model based Imputation
apache	1	0.2%	Good	Delete or Imputation

Table 6: List of variables including missing values

Visualization



Unique Values

Categorical Variables

No variable with a high proportion greater than 0.5

Numerical Variables

No variable with unique data proportion less than 5

Categorical Variable Diagnosis

Top Ranks

variables	levels	freq	ratio (%)
fate	Alive	279	61.3
fate	Dead	176	38.7
race	White	293	64.4
race	African American	130	28.6
race	Other	32	7.0
treat	Placebo	231	50.8
treat	Ibuprofen	224	49.2

Table 7: Top 10 levels of categorical variables

Numerical Variable Diagnosis

Distributions

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
id	1.00	114.50	228.00	228.00	341.50	455.00	0	0	0
apache	0.00	10.00	15.33	14.00	20.00	41.00	1	0	3
o2del	316.88	765.22	1,023.82	947.20	1,229.57	2,584.34	0	0	7
followup	1.00	290.00	532.58	720.00	720.00	720.00	0	0	0
temp0	33.10	37.50	38.01	38.17	38.67	41.67	0	0	28
temp10	33.50	36.70	37.31	37.30	37.97	40.00	0	0	5

Table 8: General list of numerical diagnosis

Zero Values

variables	min	median	max	zero	zero (%)
apache	0	14	41	1	0.2

Table 9: List of numerical diagnosis (zero)

Negative Values

No numeric variable with negative value

Outliers

List of Outliers

variables	min	median	max	outlier	outlier (%)
temp0	33.10	38.17	41.67	28	6.2
o2del	316.88	947.20	2,584.34	7	1.5
temp10	33.50	37.30	40.00	5	1.1
apache	0.00	14.00	41.00	3	0.7

Table 10: Diagnosis of numerical variable outliers

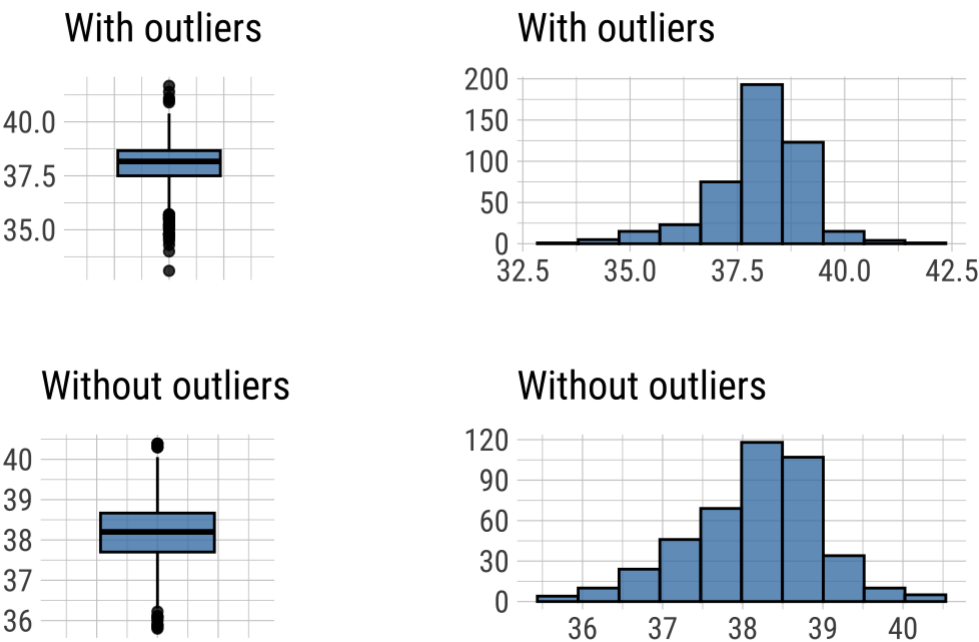
Individual Outliers

variable: temp0

Measures	Values
Outliers count	28
Outliers ratio (%)	6.15%
Mean of outliers	36.11508
Mean with outliers	38.01492
Mean without outliers	38.1395

Table 11: temp0

Outlier Diagnosis Plot (temp0)

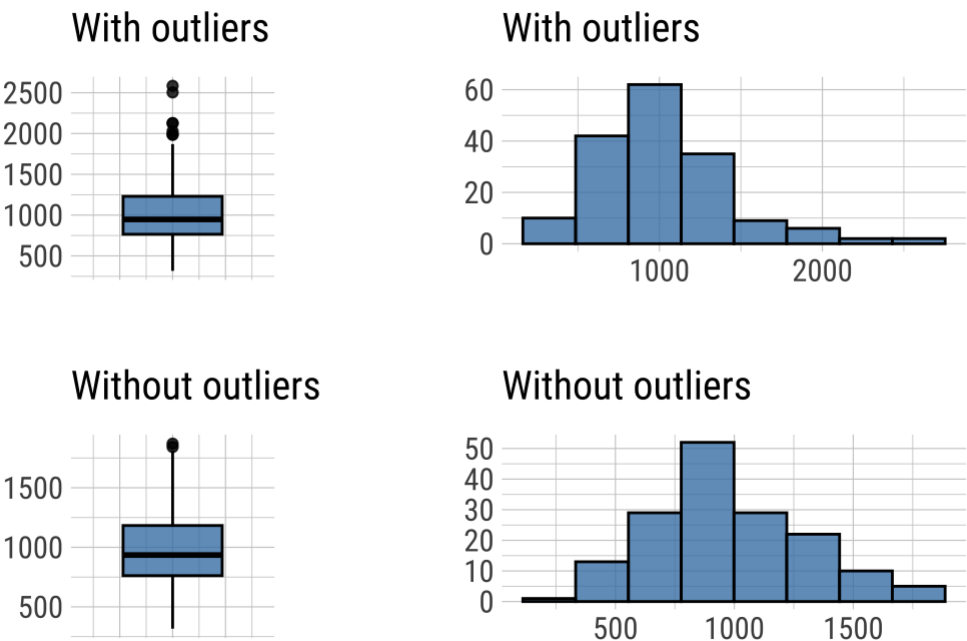


variable: o2del

Measures	Values
Outliers count	7
Outliers ratio (%)	1.54%
Mean of outliers	2191.781
Mean with outliers	1023.817
Mean without outliers	973.036

Table 11: o2del

Outlier Diagnosis Plot (o2del)



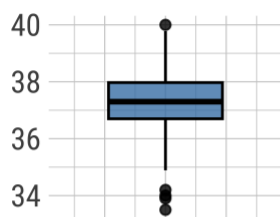
variable: temp10

Measures	Values
Outliers count	5
Outliers ratio (%)	1.1%
Mean of outliers	35.12
Mean with outliers	37.31448
Mean without outliers	37.34233

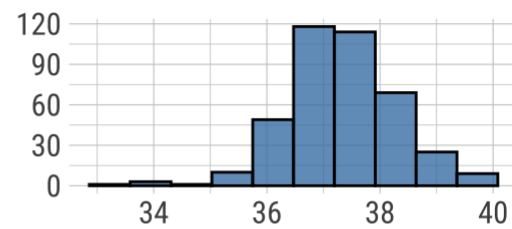
Table 11: temp10

Outlier Diagnosis Plot (temp10)

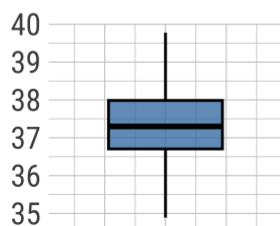
With outliers



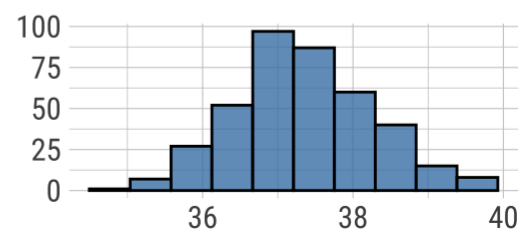
With outliers



Without outliers



Without outliers



variable: apache

Measures	Values
Outliers count	3
Outliers ratio (%)	0.66%
Mean of outliers	38
Mean with outliers	15.3304
Mean without outliers	15.1796

Table 11: apache

Outlier Diagnosis Plot (apache)

