

ML Raport

AutoPrep

January 5, 2025

Abstract

This raport has been generated with AutoPrep.

Contents

1	Overview	2
1.1	System	2
1.2	Dataset	2
2	Eda	3
2.1	Eda	3
2.2	Categorical	5
2.3	Numerical	6

1 Overview

1.1 System

System	Darwin
Machine	arm64
Processor	arm
Architecture	64bit
Python Version	3.11.10
Physical Cores	8
Logical Cores	8
CPU Frequency (MHz)	4056
Total RAM (GB)	16.00
Available RAM (GB)	5.13
Total Disk Space (GB)	460.43
Free Disk Space (GB)	251.66

Table 1: System overview.

1.2 Dataset

Number of samples	1047
Number of features	13
Number of numerical features	6
Number of categorical features	7

Table 2: Dataset Summary.

class	number of observations	Percentage
0	665	0.64
1	382	0.36

Table 3: Target class distribution.

classgit	number of observations	Percentage
pclass	0	0.00
name	0	0.00
sex	0	0.00
age	207	0.20
sibsp	0	0.00
parch	0	0.00
ticket	0	0.00
fare	1	0.00
cabin	813	0.78
embarked	1	0.00
boat	672	0.64
body	948	0.91
home__dest	453	0.43

Table 4: Missing values distribution.

class	type	dtype	space usage
pclass	numerical	int64	16.8 kB
name	categorical	object	96.4 kB
sex	categorical	category	9.7 kB
age	numerical	float64	16.8 kB
sibsp	numerical	int64	16.8 kB
parch	numerical	int64	16.8 kB
ticket	categorical	object	75.1 kB
fare	numerical	float64	16.8 kB
cabin	categorical	object	48.6 kB
embarked	categorical	category	9.7 kB
boat	categorical	object	51.8 kB
body	numerical	float64	16.8 kB
home__dest	categorical	object	68.2 kB

Table 5: Features description.

2 Eda

2.1 Eda

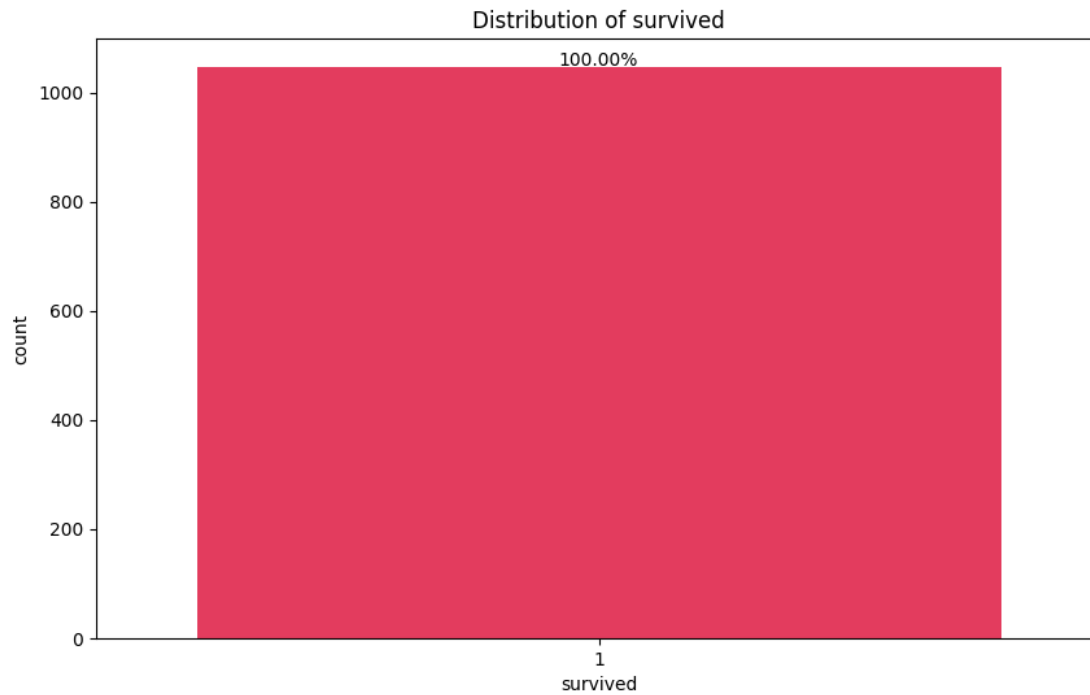


Figure 1: Target distribution.

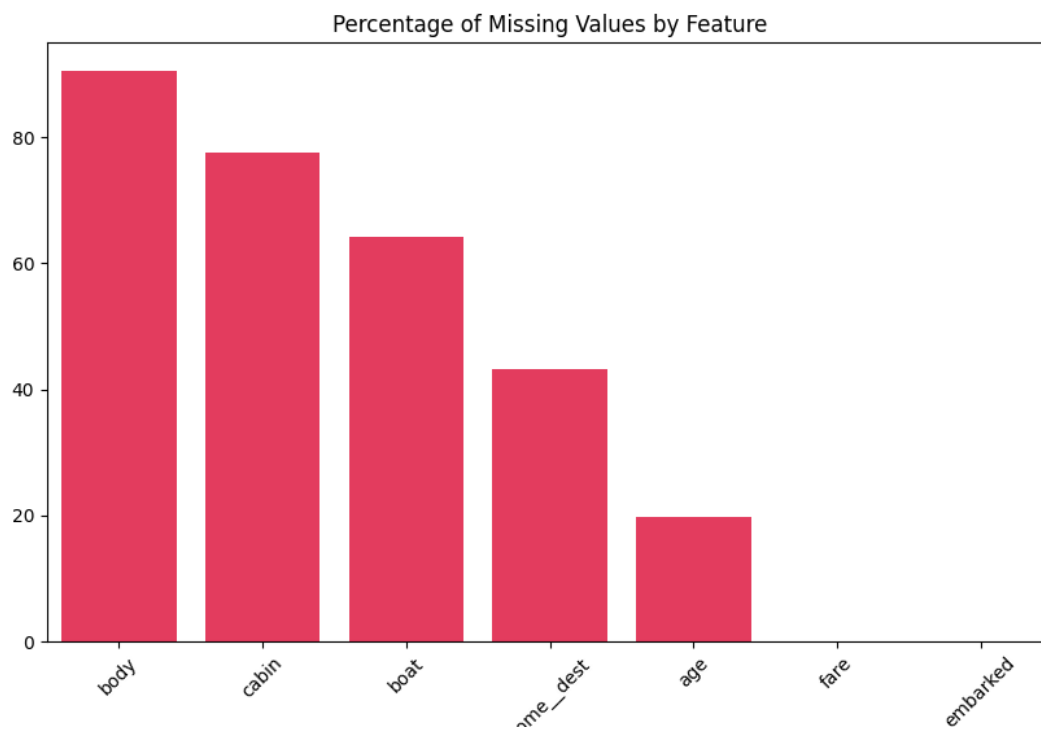


Figure 2: Missing values.

2.2 Categorical

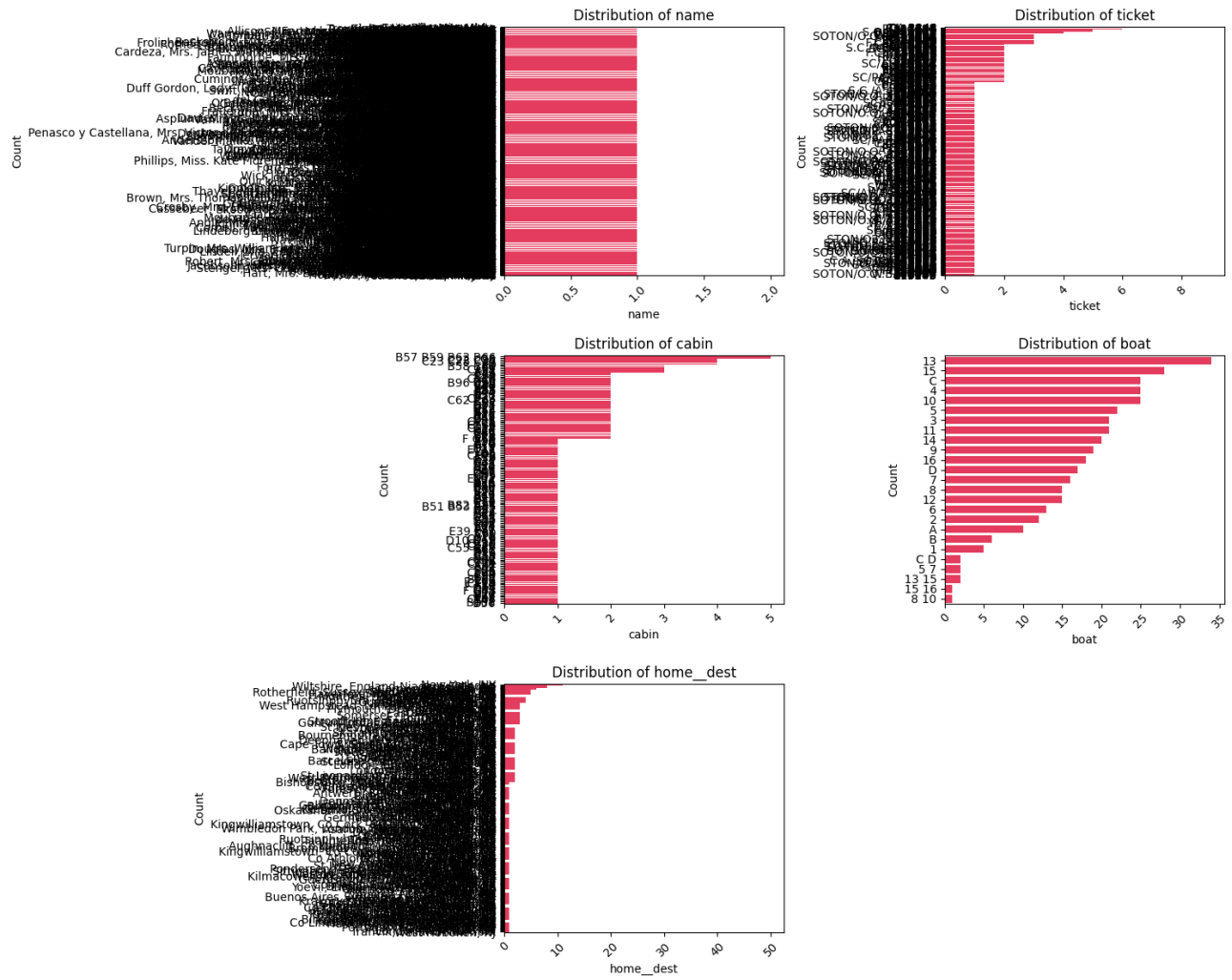


Figure 3: Categorical distribution.

2.3 Numerical

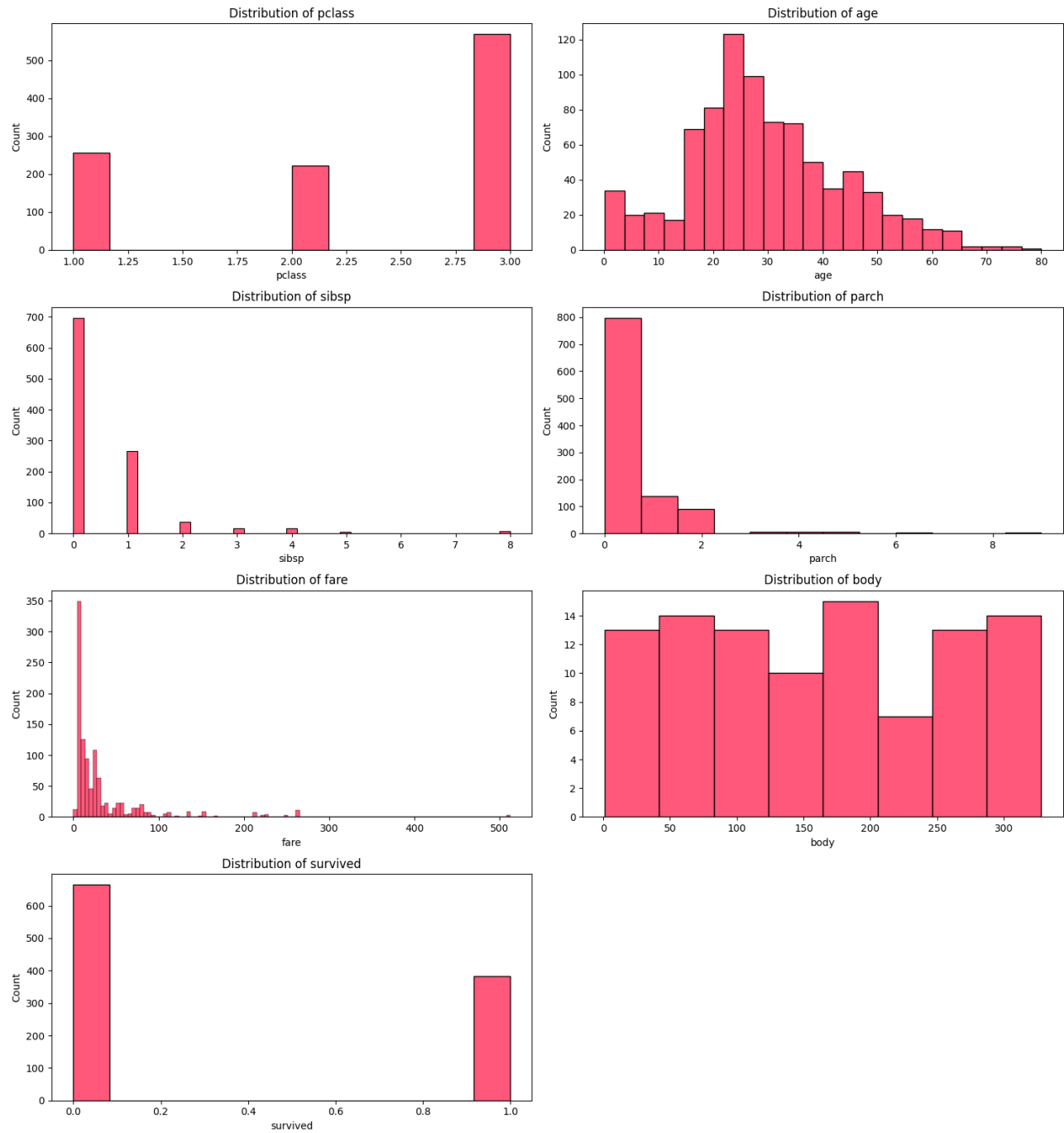


Figure 4: Numerical distribution.

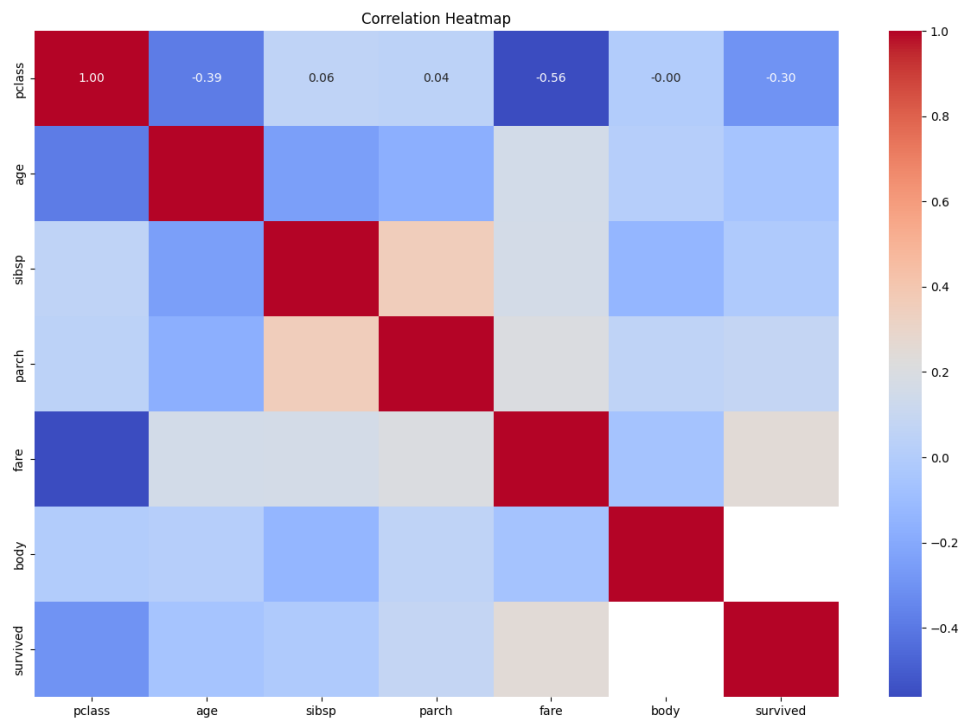


Figure 5: Correlation heatmap.

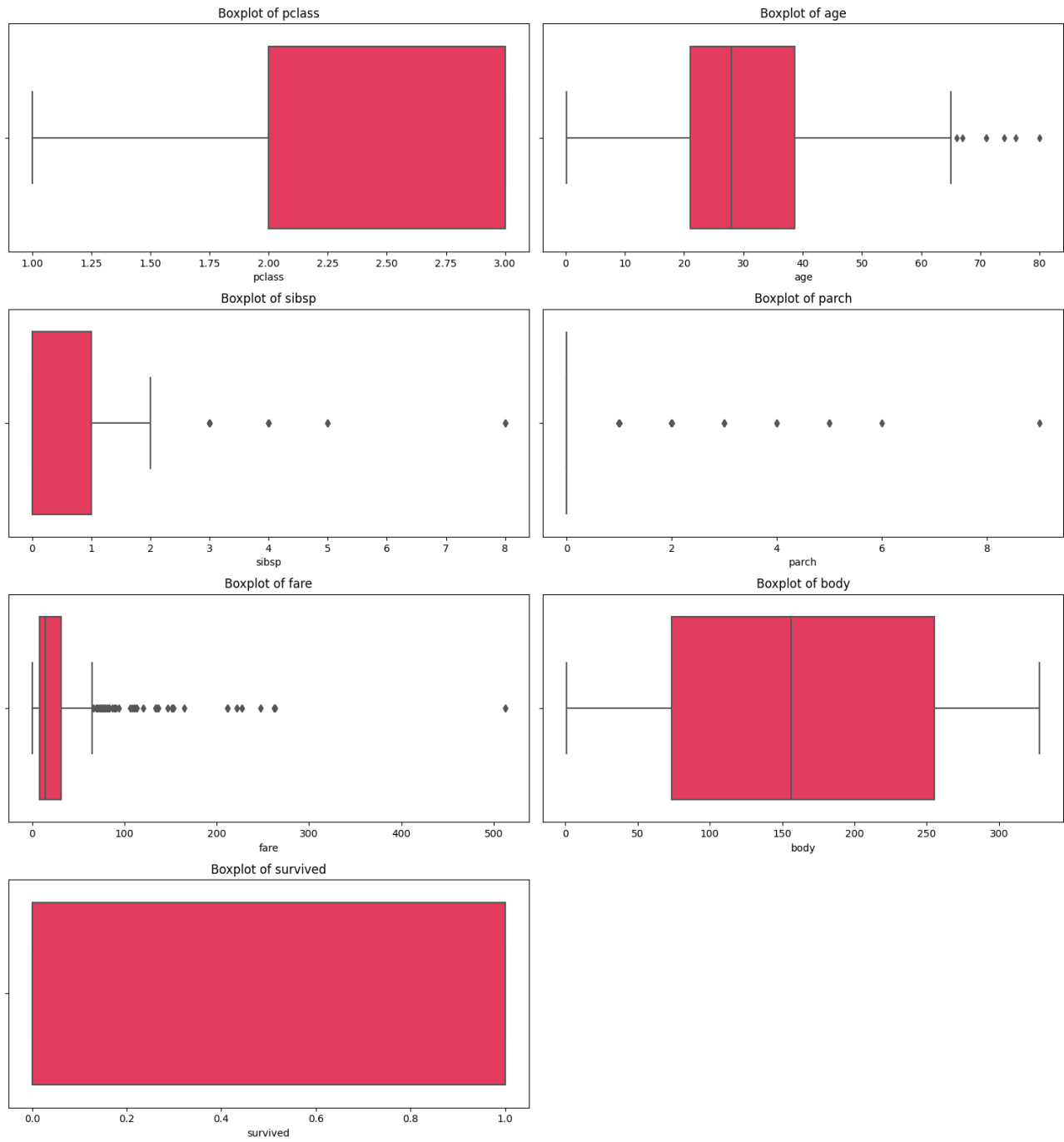


Figure 6: Boxplot.