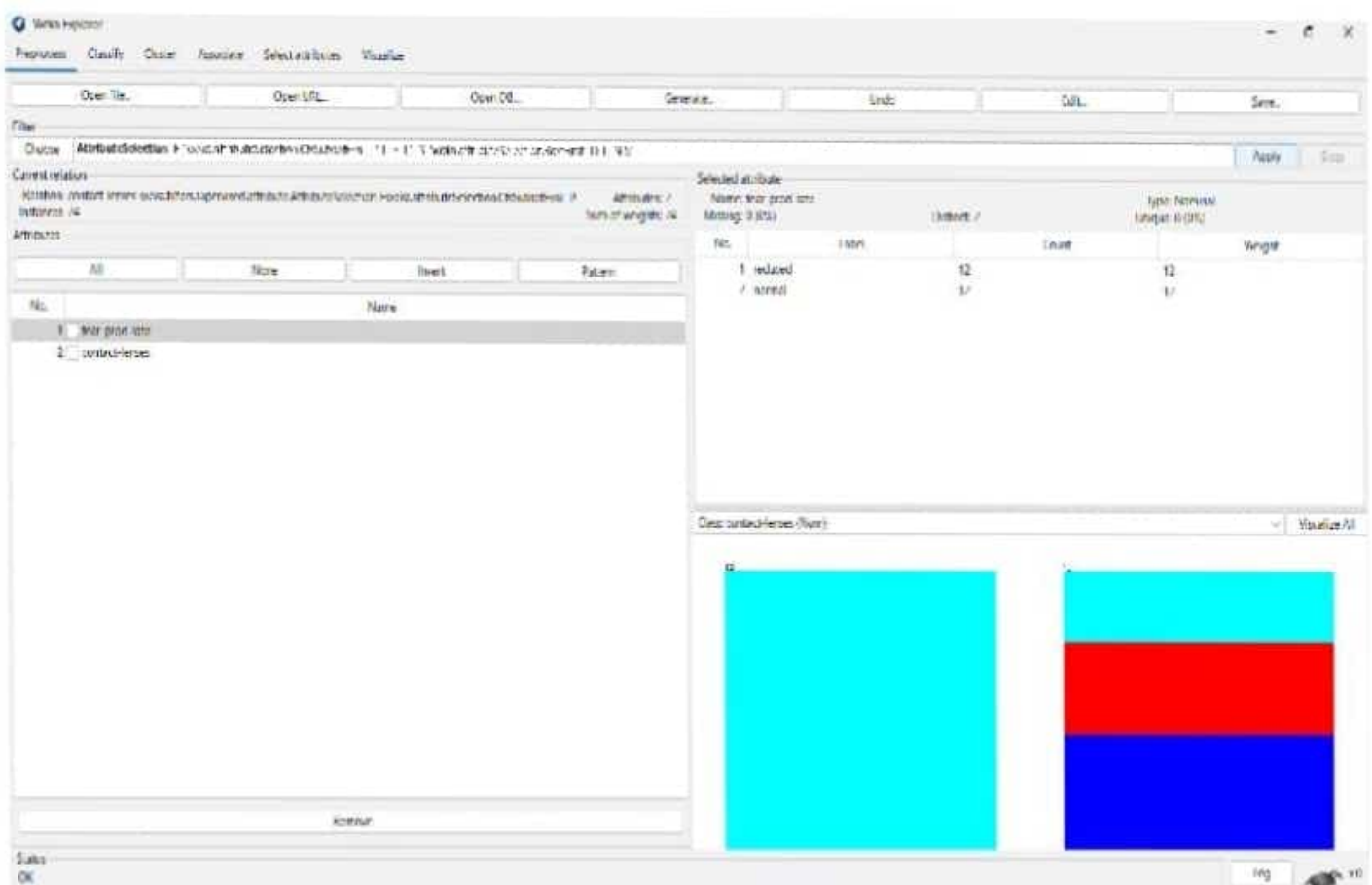
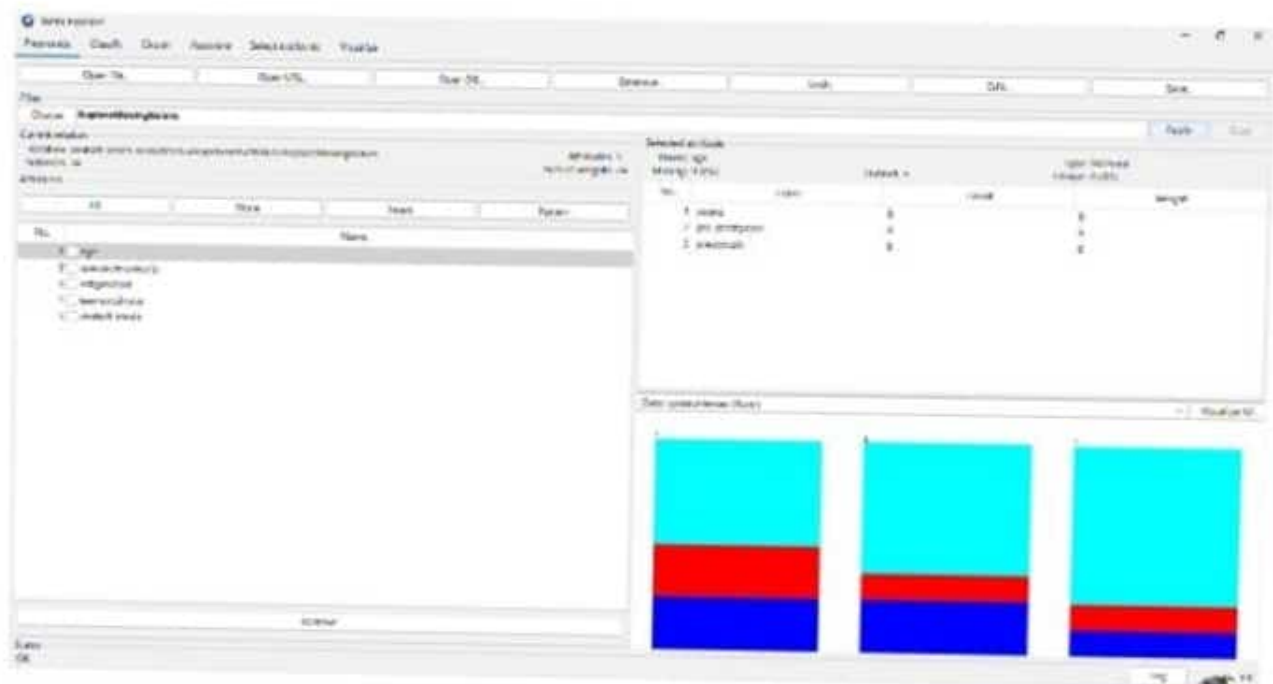


## Datapreprocessing:

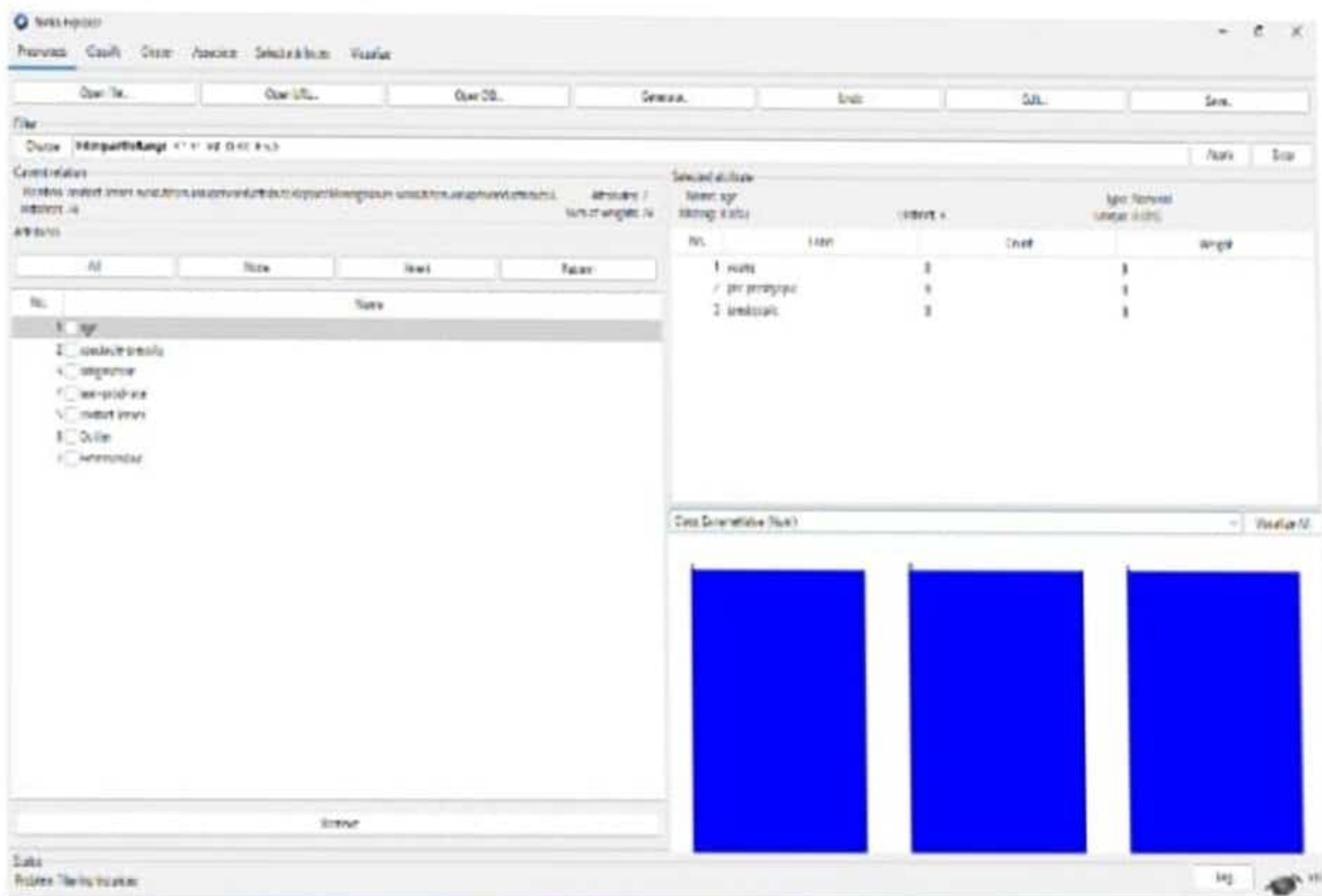
### Datapreprocessing:



Remove Missing values:



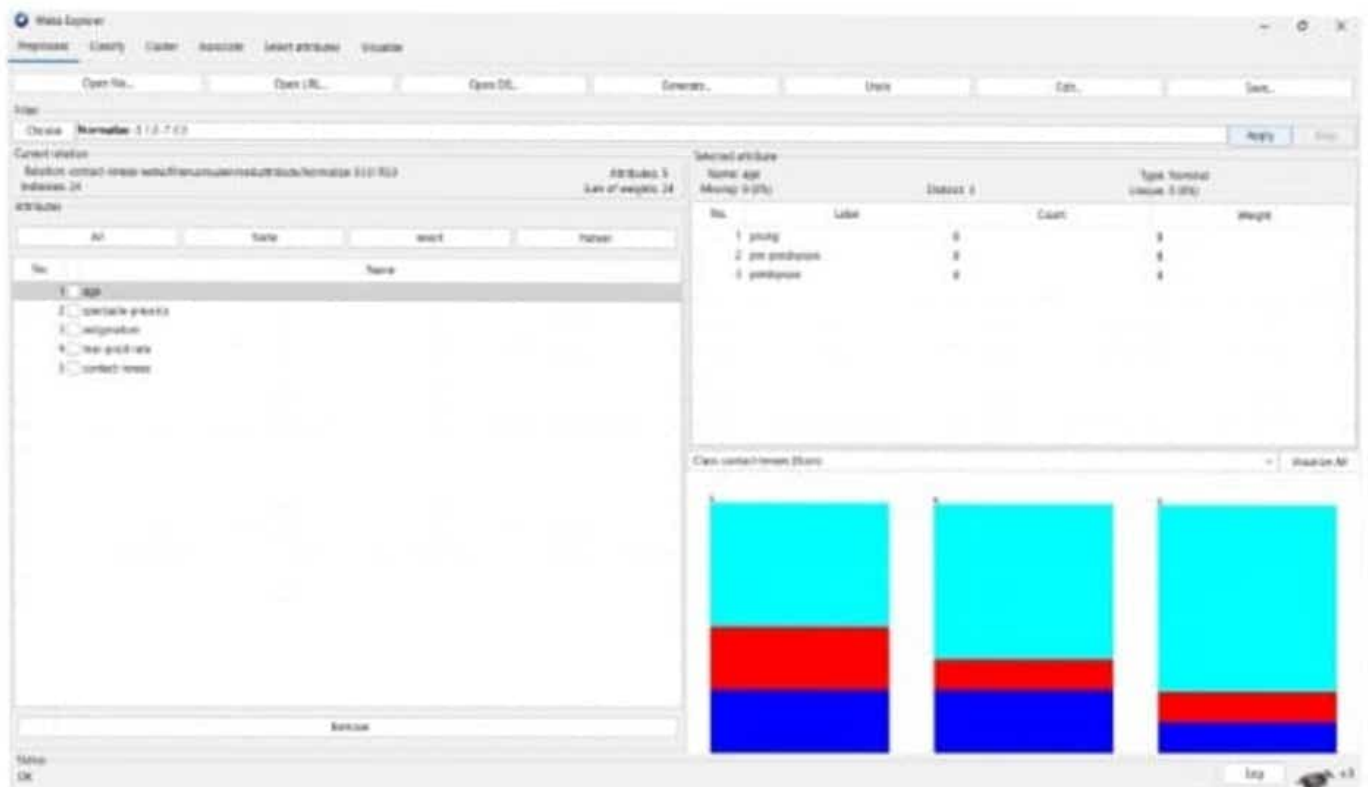
Interquartile range:



### Classification(Bayes):

The screenshot shows the 'New Report' dialog box in SAP. The 'Navigation' tab is selected. The 'Table' field is set to 'ZTR007 - Lernaktivitäten'. The 'Table' field is highlighted in blue. The 'Table' field is set to 'ZTR007 - Lernaktivitäten'. The 'Table' field is highlighted in blue.

# Normalize:



Decision tree classification:



# Clustering:

## K-means:

The screenshot shows the Weka Explorer application with the K-means clustering algorithm applied to the 'breast-wdbc' dataset. The 'Cluster mode' is set to 'Use training set'. The 'Cluster output' pane displays the following information:

- Number of iterations: 3
- Within cluster sum of squared errors: 47.0
- Initial starting points (random)
- Cluster 0: per-pendpops,spen,se,small,wtb
- Cluster 1: per-pendpops,spen,se,small,wtb
- Missing values properly replaced with mean/med
- Final cluster assignment table:

Attribute	Cluster		
	0 (24.1%)	1 (22.1%)	2 (22.1%)
age	young	young	young
specchic-possibp	spice	spice	spice
sechpense	on	on	on
near-pool-vote	induced	induced	induced
constant-classes	none	none	none

Time taken to build model (full training data): 1.3 seconds

Model and evaluation on training set: OK

Clustered Statistics:

Cluster	Count	Percentage
0	12	100%
1	12	100%



# EM :

Weka Explorer

Preprocessor   Classify   Cluster   Associate   Select attributes   Visualize

Cluster: **EM** (100 % - 1.813 sec) (4 x 100 x 4 x 100 x 100) x M100 x 4 x 10 num dist 1.0100

Cluster mode

- ☒ Use training set
- ☐ Supplied test set
- ☐ Percentage split
- ☐ Cases to cluster evaluation
- ☐ Display cluster centers
- ☒ Store clusters for visualization

Ignore attributes

Start   Stop

Need not right-click for options

21.04.23 - Supermarket

21.04.23 - EM

Cluster output

Attributes	0	1
Instances	20,422	19,286
Age		
young	0.1240	0.0990
young-adult	0.2045	0.1338
middle-aged	0.5089	0.3517
(total)	01.0000	01.0000
specific-to-previous		
yes	0.0000	0.0000
no	0.0000	0.0000
(total)	00.0000	00.0000
anticipation		
no	0.0000	0.0000
yes	0.0000	0.0000
(total)	00.0000	00.0000
new-product		
yes	0.0000	0.0000
no	0.0000	0.0000
(total)	00.0000	00.0000
customer-loyalty		
no	0.0000	0.0000
yes	0.0000	0.0000
(total)	00.0000	00.0000
new-product		
yes	0.0000	0.0000
no	0.0000	0.0000
(total)	00.0000	00.0000
customer-loyalty		
no	0.0000	0.0000
yes	0.0000	0.0000
(total)	00.0000	00.0000

Time taken to build model: 17.01 seconds (total training data: 1.0100)

\*\*\* Model and evaluation on training set \*\*\*

Clustered Instances

0	1
15.1000	14.1860
0.0000	0.0000

Log Likelihood: -5.00000

Status: OK

142

