



# Eye-Tracking Data Preprocessing & Recommendations

Upload your CSV file

 Drag and drop file here  
Limit 200MB per file • CSV

E

 Frage 1\_001\_Anonymous 09-05-23 12h04m.csv 1.6MB

File successfully uploaded.

## Data Preview

	Row	Timestamp	EventSource	SlideEvent	StimType	Duration	CollectionPhase	SourceStimuliName	EventSource.1	CPU Sys	Memory Sys	CPU Proc	Memory Proc	EventSource.2	
0	1	30.9133	1	StartSlide	TestImage	90,000	StimuliDisplay	Frage 1	None	None	None	None	None	None	
1	2	32.508	None	None	None	None	None	Frage 1	None	None	None	None	None	1	
2	3	49.1635	None	None	None	None	None	Frage 1	None	None	None	None	None	1	
3	4	65.8267	None	None	None	None	None	Frage 1	None	None	None	None	None	1	
4	5	82.474	None	None	None	None	None	Frage 1	None	None	None	None	None	1	

## Select Features to Analyze

Choose numerical features

Gaze X

X

Select Feature for Preview

Gaze X

## Feature Diagnostics

Missing Rate  
26.9%

Outlier Rate (IQR)  
2.2%

Skewness  
1.01

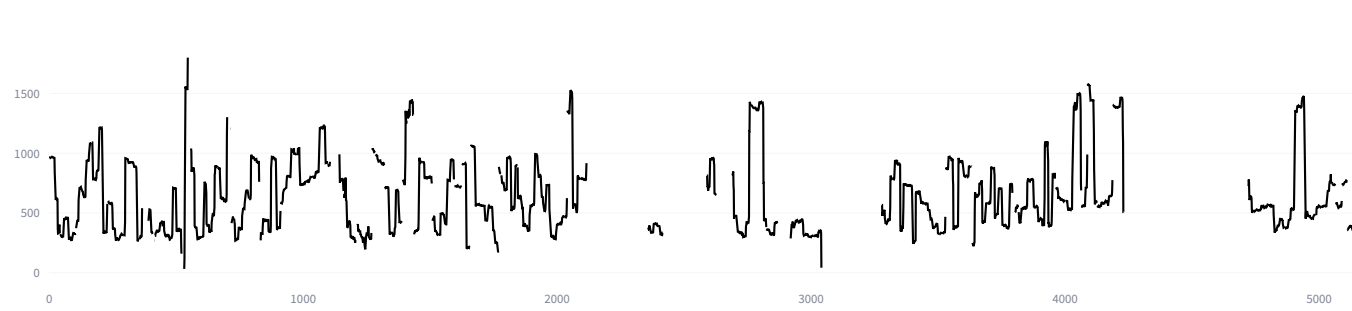
Kurtosis  
0.49

Pipeline heuristic based on missing: 26.9%, outliers: 2.2%.

- ♦ **Mixed case:** • Missing: KNN or MICE depending on complexity. • Outliers: Isolation Forest or MAD (depending on skewness). • Scaling: Min-Max for moderate outliers, otherwise RobustScaler

detailed recommendations in each section as guidance.

## Original Feature Plot



## Missing Value Strategy Preview

Select Missing Value Methods to Compare

LOCF

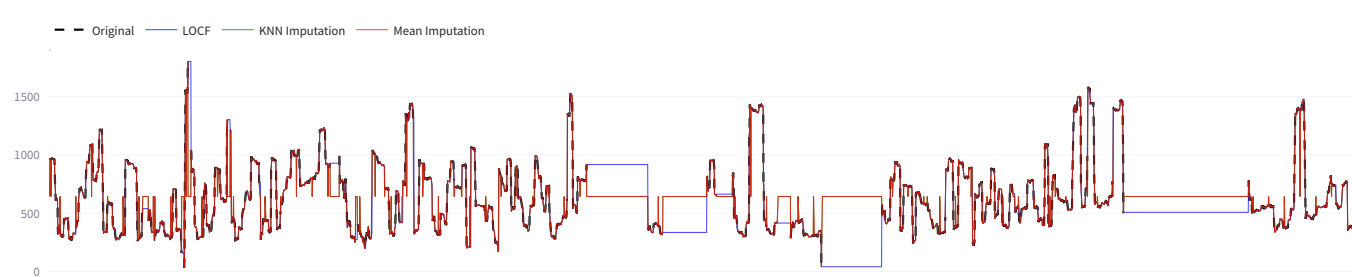
X

KNN Imputation

X

Mean Imputation

X





Recommendation (Missing Values):

Missing rate: 26.9%.

♦ Recommended:

KNN or MICE, but overall data quality is questionable.

♦ Alternatives:

• Drop the feature if missingness is extremely high (e.g. >40%).

• Check whether a sensor/channel systematically (design issue rather than preprocessing).

☒

Apply this MV method for next steps

Mean Imputation

Preview: Outlier Detection

Select Outlier Methods to Compare

IQR

MAD

Isolation Forest

☒

Apply this Outlier Method

IQR

Z-Score threshold



Isolation Forest contamination



Recommendation (Outliers):

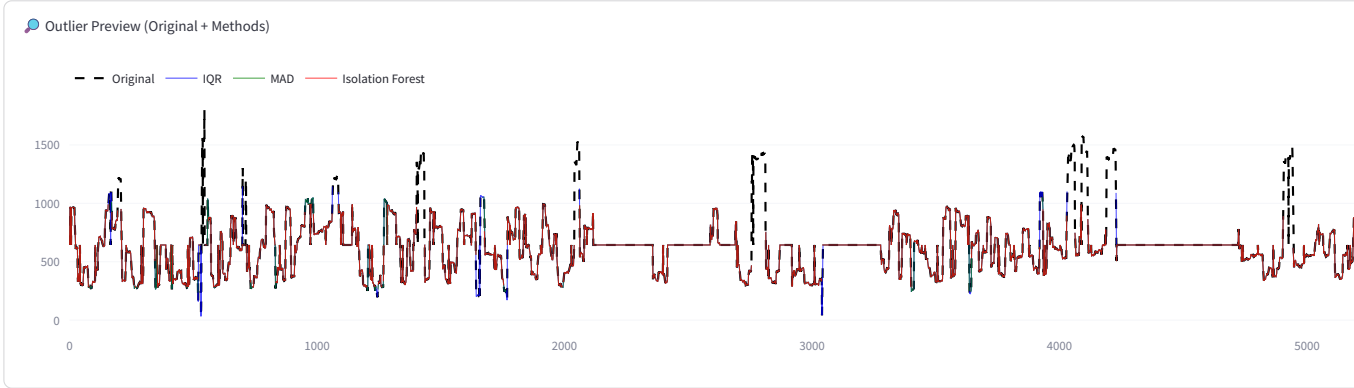
Estimated outlier rate (IQR-based): 2.2%. Skewness: 1.01.

♦ Recommended:

MAD or Isolation Forest (robust under strong skew).

♦ Alternatives:

Winsorization for visualization; avoid Z-Score here.



Preview: Normalization

Manual Y-Axis Zoom

Y-Axis Min (optional)

-1,00

Y-Axis Max (optional)

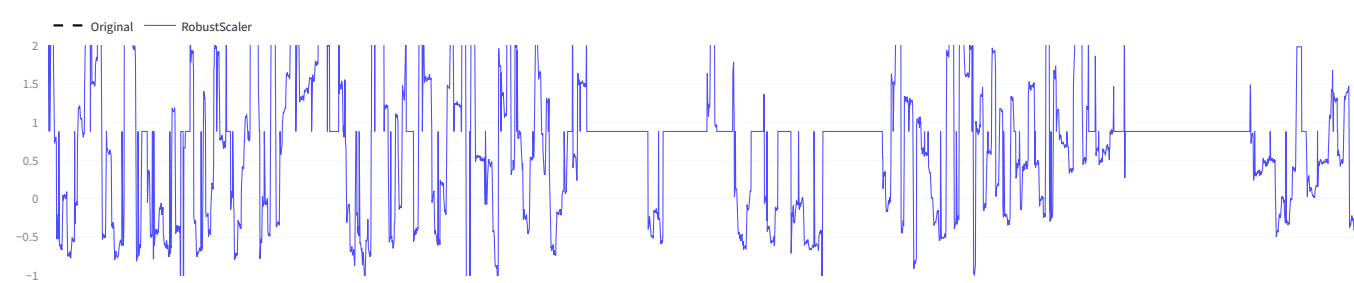
2,00

☒

Use custom Y-axis range

Select normalization methods to preview

RobustScaler





Recommendation (Normalization):

Skewness: 1.01, outlier rate: 2.2%.

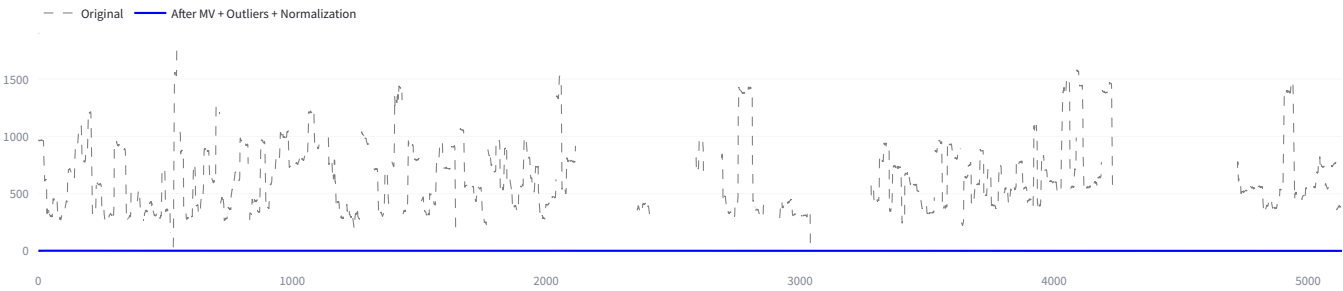
♦ Recommended: RobustScaler (robust to outliers and skew, often good for pupil features). ♦ Alternatives: • Log/power transform + StandardScaler for heavy right skew. • Min-Max only with aggressive outlier handling.

Apply this Normalization Method

RobustScaler

Applied Preprocessing Summary (before optional smoothing)

- Missing value method: Mean Imputation
- Outlier method: IQR
- Normalization: RobustScaler



Optional Post-Normalization Steps

These steps act on the applied normalization (blue line above). They are primarily for visual smoothing and are not included in the CSV export.

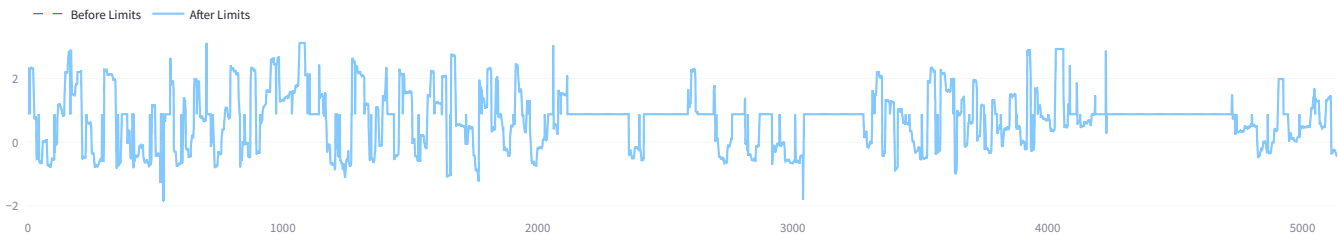
Apply manual feature limits

Lower limit

-1,86

Upper limit

3,16



Apply Butterworth low-pass filter

Cutoff Frequency

0.10

0.01

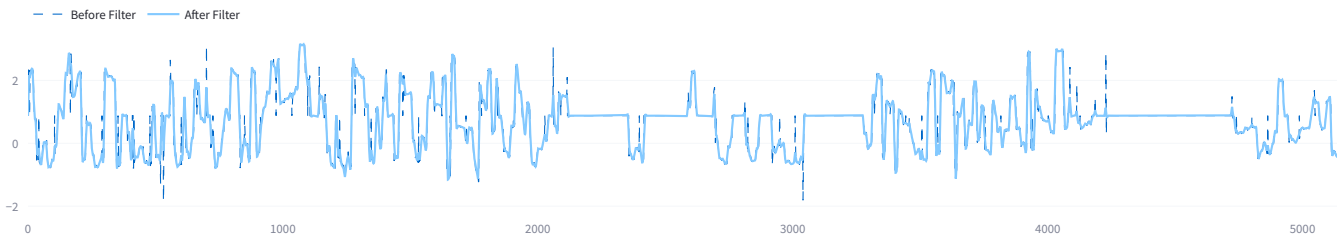
Sampling Frequency (Hz)

1,00

Filter Order

2

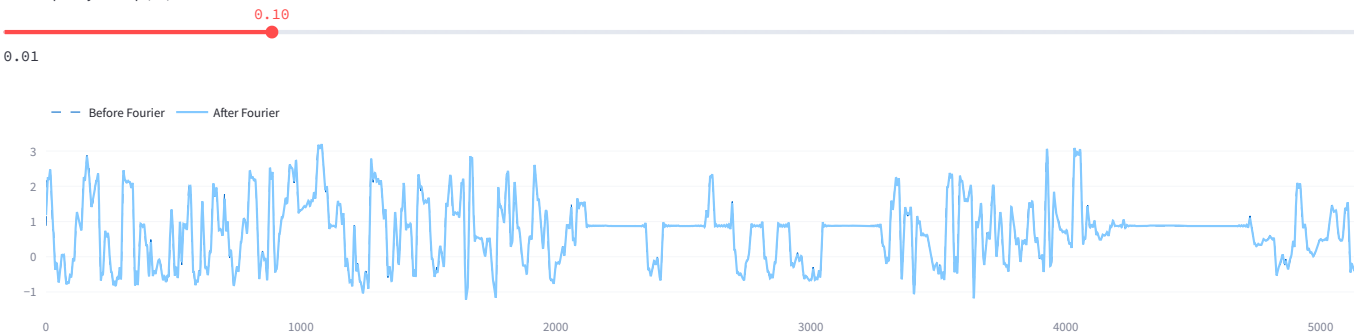
1





☒ Apply Fourier Smoothing

Max frequency to keep (Hz)



Show Final Comparison

Summary Statistics Before vs After

	Metric	Before	
0	Mean	643.4557	
1	Std	305.1007	
2	Missing %	26.9147	
3	Outlier Removed %	0.0000	

Final Data Preview

	Row	Timestamp	EventSource	SlideEvent	StimType	Duration	CollectionPhase	SourceStimuliName	EventSource.1	CPU Sys	Memory Sys	CPU Proc	Memory Proc	EventSource.2
0	1	30.9133	1	StartSlide	TestImage	90,000	StimuliDisplay	Frage 1	None	None	None	None	None	None
1	2	32.508	None	None	None	None	None	Frage 1	None	None	None	None	None	1
2	3	49.1635	None	None	None	None	None	Frage 1	None	None	None	None	None	1
3	4	65.8267	None	None	None	None	None	Frage 1	None	None	None	None	None	1
4	5	82.474	None	None	None	None	None	Frage 1	None	None	None	None	None	1

Download Processed CSV