# Dataset Comparison Report

## Comparison Summary

X_train_processed.csv: Comparison Completed

X_test_processed.csv: Comparison Completed

X_train_with_new_features.csv: Comparison Completed

## Column Changes

Added Columns: chas_1

  NaN

Removed Columns: medv

chas

Added Columns: chas_1

  NaN

Removed Columns: medv

chas

Added Columns: tax_to_rm

 lstat_sq

 log_crim

  chas_1

# Dataset Comparison Report

Removed Columns: medv

chas

 NaN

 NaN

## Numerical Differences

count: -102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

mean:   -3.613524

 -11.363636

 -11.136779

 -0.554695

 -6.284634

# Dataset Comparison Report

-68.574901

-3.795043

-9.549407

-408.237154

-18.455534

-356.674032

-12.653063

std:   -7.600305

-22.321213

-5.859113

0.885362

0.298623

-27.147621

-1.10447

-7.706019

-167.535876

-1.163706

-90.293624

-6.139822

min:   -0.412461

-0.50032

-1.947791

-1.843711

# Dataset Comparison Report

-7.022733

-5.248206

-2.387991

-1.974048

-188.307684

-15.168944

-4.212249

-3.240559

25%:   -0.480021

-0.50032

-6.040273

-1.337783

-6.485849

-45.847826

-2.908373

-4.62436

-279.753356

-18.082101

-375.16762

-7.750506

50%:   -0.634185

-0.50032

-10.042167

# Dataset Comparison Report

-0.695233

-6.357943

-77.17298

-3.50457

-5.507797

-330.446066

-18.87853

-391.056999

-11.575776

75%:   -3.722904

-12.135411

-17.066763

0.009861

-6.170676

-93.177527

-4.440227

-23.691859

-664.421566

-19.354657

-395.79334

-16.403693

max:  -79.345511

-96.175773

# Dataset Comparison Report

-25.306837

1.804394

-5.302436

-98.875419

-8.2186

-22.293109

-709.150427

-20.34601

-396.45582

-34.377466

count: -404.0

-404.0

-404.0

-404.0

-404.0

-404.0

-404.0

-404.0

-404.0

-404.0

-404.0

-404.0

mean:   -3.611062

# Dataset Comparison Report

-11.407759

-11.027472

-0.630192

-6.503465

-68.571625

-3.825695

-9.437823

-408.111467

-18.149729

-356.652568

-12.516349


std:   -7.759989

-22.280962

-5.884097

0.807936

0.234231

-27.115578

-1.16107

-7.639326

-167.467852

-1.356039

-90.303735

-6.120922

# Dataset Comparison Report

min:    -0.41277

   -0.50032

   -1.988453

   -1.784166

   -7.448942

   -5.130181

   -2.369434

   -1.974048

-188.301658

  -14.989244

   -4.142705

   -3.078623


25%:   -0.479352

   -0.50032

   -5.899046

   -1.369682

   -6.514427

 -45.910415

   -2.825654

   -4.62436

-279.736787

  -17.363304

-375.172568

   -7.621661

# Dataset Comparison Report

50%:   -0.639995

   -0.50032

   -9.747369

   -0.746272

   -6.413689

   -77.280275

   -3.421111

   -5.507797

   -330.355687

   -18.653905

   -391.047158

   -11.377229


75%:   -3.574043

   -13.00032

   -17.066763

   -0.07095

   -6.392954

   -93.121197

   -4.739005

   -22.293109

   -664.421566

   -19.354657

   -395.78082

   -16.172054

# Dataset Comparison Report

max: -84.222516

-96.392

-25.306837

1.804394

-6.355256

-98.875419

-8.883924

-22.293109

-709.150427

-20.34601

-396.45582

-34.516872

count: -102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

-102.0

# Dataset Comparison Report

mean:  -3.613524

 -11.363636

 -11.136779

 -0.554695

 -6.284634

-68.574901

 -3.795043

 -9.549407

-408.237154

 -18.455534

-356.674032

 -12.653063


std:  -7.600305

 -22.321213

 -5.859113

  0.885362

  0.298623

-27.147621

 -1.10447

 -7.706019

-167.535876

 -1.163706

-90.293624

 -6.139822

# Dataset Comparison Report

min:   -0.412461

-0.50032

-1.947791

-1.843711

-7.022733

-5.248206

-2.387991

-1.974048

-188.307684

-15.168944

-4.212249

-3.240559

25%:   -0.480021

-0.50032

-6.040273

-1.337783

-6.485849

-45.847826

-2.908373

-4.62436

-279.753356

-18.082101

-375.16762

-7.750506

# Dataset Comparison Report

50%:  -0.634185

 -0.50032

-10.042167

-0.695233

-6.357943

-77.17298

-3.50457

-5.507797

-330.446066

-18.87853

-391.056999

-11.575776


75%:  -3.722904

-12.135411

-17.066763

 0.009861

-6.170676

-93.177527

-4.440227

-23.691859

-664.421566

-19.354657

-395.79334

-16.403693

# Dataset Comparison Report

max: -79.345511

-96.175773

-25.306837

1.804394

-5.302436

-98.875419
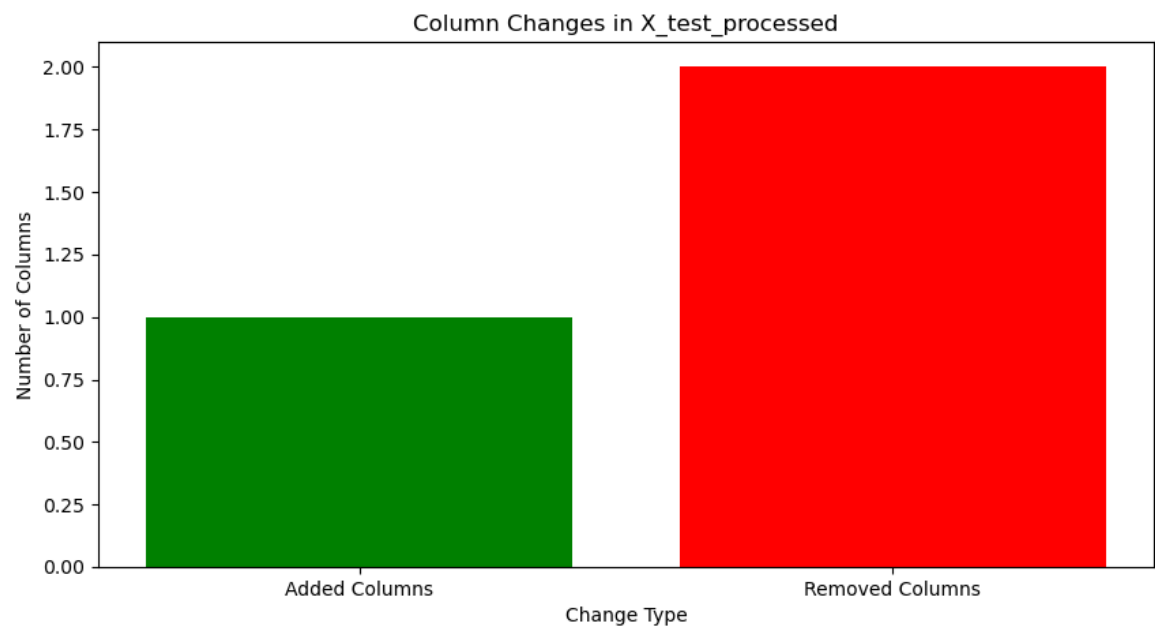
-8.2186

-22.293109

-709.150427
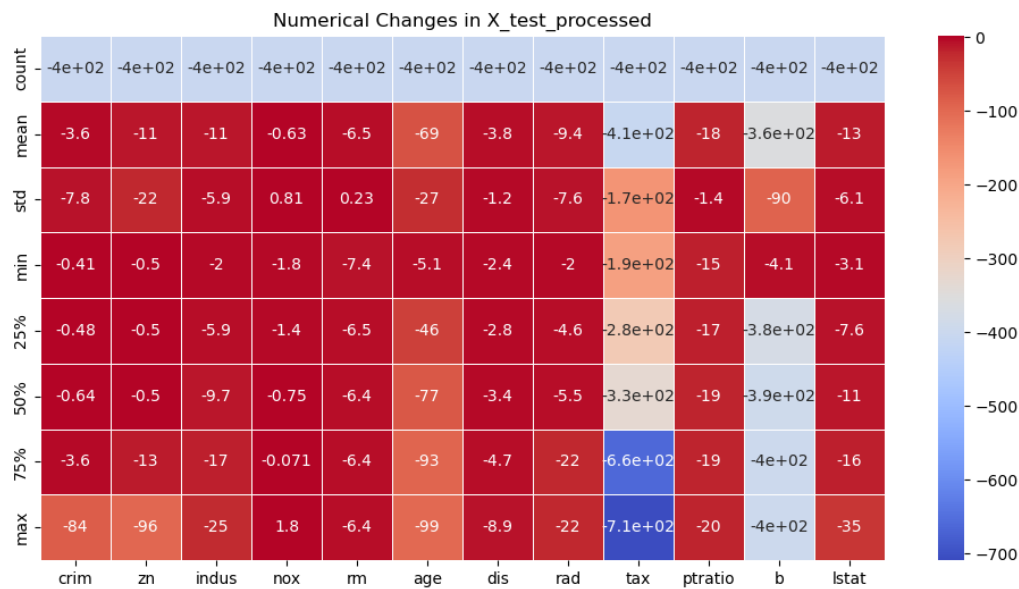
-20.34601

-396.45582

-34.377466

## Visual Comparisons

## X test processed column changes

# Dataset Comparison Report

## X test processed numerical changes



Numerical Changes in X_test_processed

| | crim | zn | indus | nox | rm | age | dis | rad | tax | ptratio | b | lstat |
|------|------|------|-------|------|------|------|------|------|---------|---------|---------|------|
| count | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 | -4e+02 |
| mean | -3.6 | -11 | -11 | -0.63 | -6.5 | -69 | -3.8 | -9.4 | -4.1e+02 | -18 | -3.6e+02 | -13 |
| std | -7.8 | -22 | -5.9 | 0.81 | 0.23 | -27 | -1.2 | -7.6 | -1.7e+02 | -1.4 | -90 | -6.1 |
| min | -0.41 | -0.5 | -2 | -1.8 | -7.4 | -5.1 | -2.4 | -2 | -1.9e+02 | -15 | -4.1 | -3.1 |
| 25% | -0.48 | -0.5 | -5.9 | -1.4 | -6.5 | -46 | -2.8 | -4.6 | -2.8e+02 | -17 | -3.8e+02 | -7.6 |
| 50% | -0.64 | -0.5 | -9.7 | -0.75 | -6.4 | -77 | -3.4 | -5.5 | -3.3e+02 | -19 | -3.9e+02 | -11 |
| 75% | -3.6 | -13 | -17 | -0.071 | -6.4 | -93 | -4.7 | -22 | -6.6e+02 | -19 | -4e+02 | -16 |
| max | -84 | -96 | -25 | 1.8 | -6.4 | -99 | -8.9 | -22 | -7.1e+02 | -20 | -4e+02 | -35 |

## X train processed column changes



Column Changes in X_train_processed

## X train processed numerical changes

# Dataset Comparison Report



Numerical Changes in X_train_processed
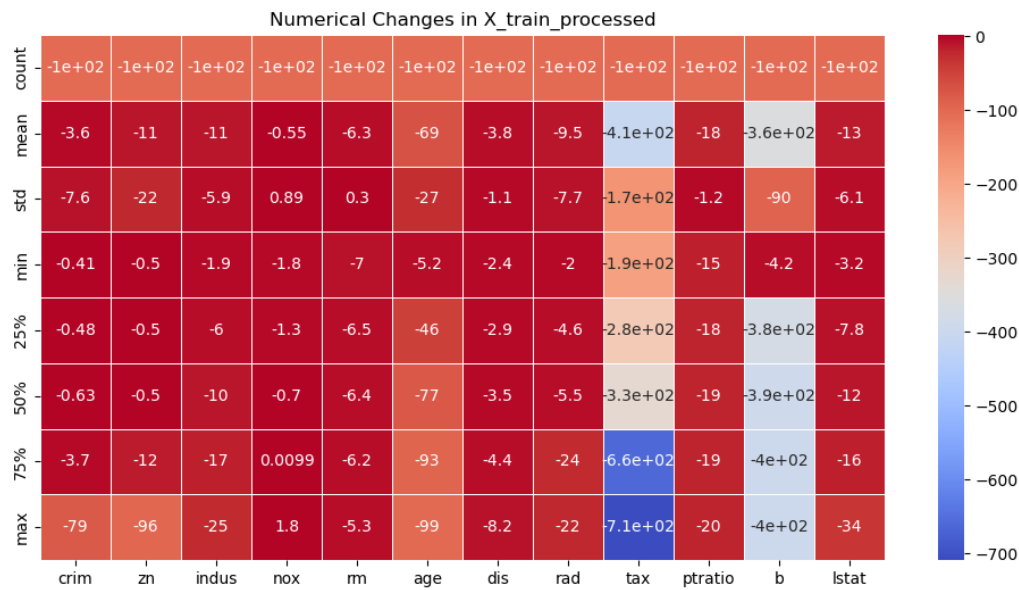
**X train with new features column changes**



Column Changes in X_train_with_new_features
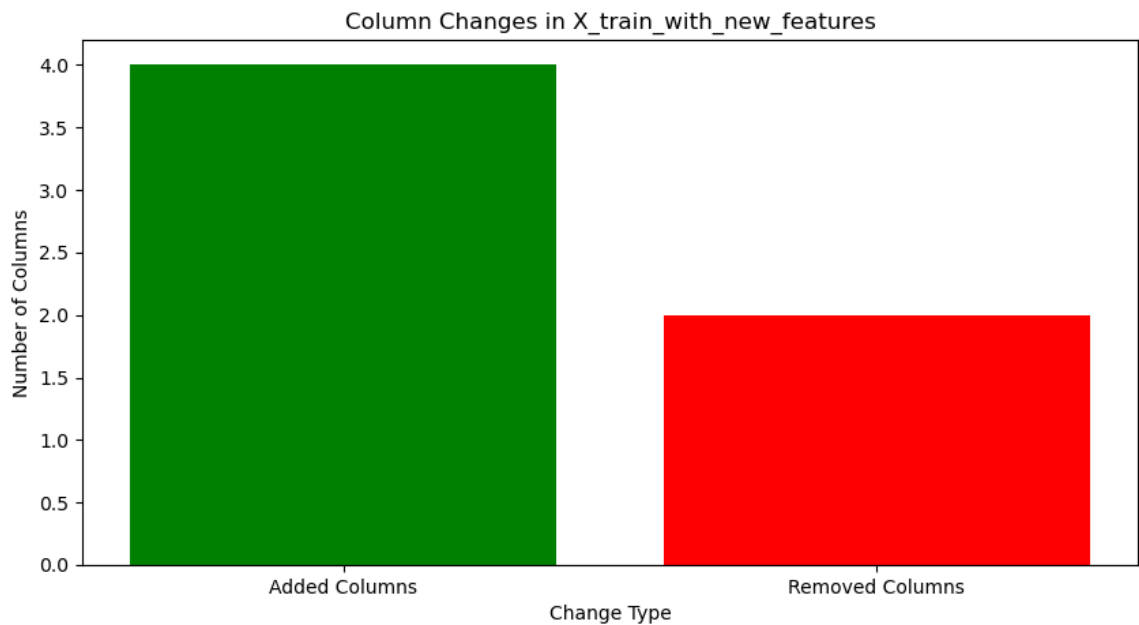
**X train with new features numerical changes**

# Dataset Comparison Report

## Numerical Changes in X_train_with_new_features



| | crim | zn | indus | nox | rm | age | dis | rad | tax | ptratio | b | lstat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 | -1e+02 |
| **mean** | -3.6 | -11 | -11 | -0.55 | -6.3 | -69 | -3.8 | -9.5 | -4.1e+02 | -18 | -3.6e+02 | -13 |
| **std** | -7.6 | -22 | -5.9 | 0.89 | 0.3 | -27 | -1.1 | -7.7 | -1.7e+02 | -1.2 | -90 | -6.1 |
| **min** | -0.41 | -0.5 | -1.9 | -1.8 | -7 | -5.2 | -2.4 | -2 | -1.9e+02 | -15 | -4.2 | -3.2 |
| **25%** | -0.48 | -0.5 | -6 | -1.3 | -6.5 | -46 | -2.9 | -4.6 | -2.8e+02 | -18 | -3.8e+02 | -7.8 |
| **50%** | -0.63 | -0.5 | -10 | -0.7 | -6.4 | -77 | -3.5 | -5.5 | -3.3e+02 | -19 | -3.9e+02 | -12 |
| **75%** | -3.7 | -12 | -17 | 0.0099 | -6.2 | -93 | -4.4 | -24 | -6.6e+02 | -19 | -4e+02 | -16 |
| **max** | -79 | -96 | -25 | 1.8 | -5.3 | -99 | -8.2 | -22 | -7.1e+02 | -20 | -4e+02 | -34 |