

# Data Science Lab 5

<b>Team nr:</b> 2	<b>Student 1:</b> Åmund Grimstad	<b>IST nr:</b> 1116675
	<b>Student 2:</b> Arthur de Arruda Chau	<b>IST nr:</b> 1116090
	<b>Student 3:</b> Benjamin Raymond Kuhn	<b>IST nr:</b> 1115778
	<b>Student 4:</b> João Rafael Freitas Lourenço	<b>IST nr:</b> 425699

## TIME SERIES ANALYSIS

### Traffic Time Series

#### DATA PROFILING

*Data Dimensionality and Granularity*

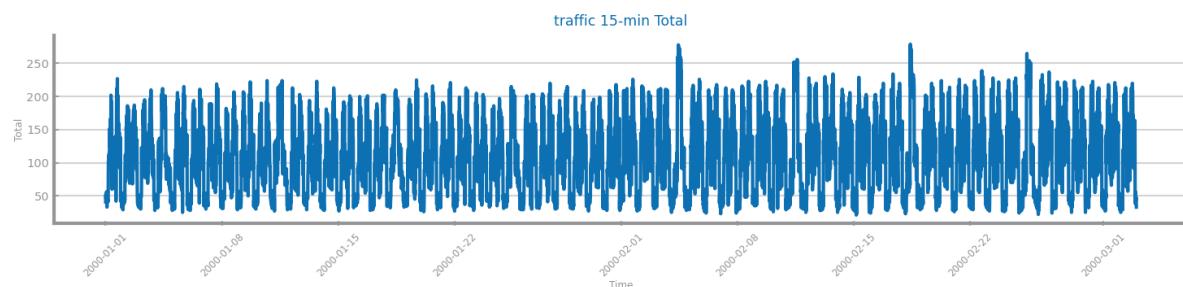


Figure 1: Traffic Time Series dimensionality over time

### traffic Total aggregation study

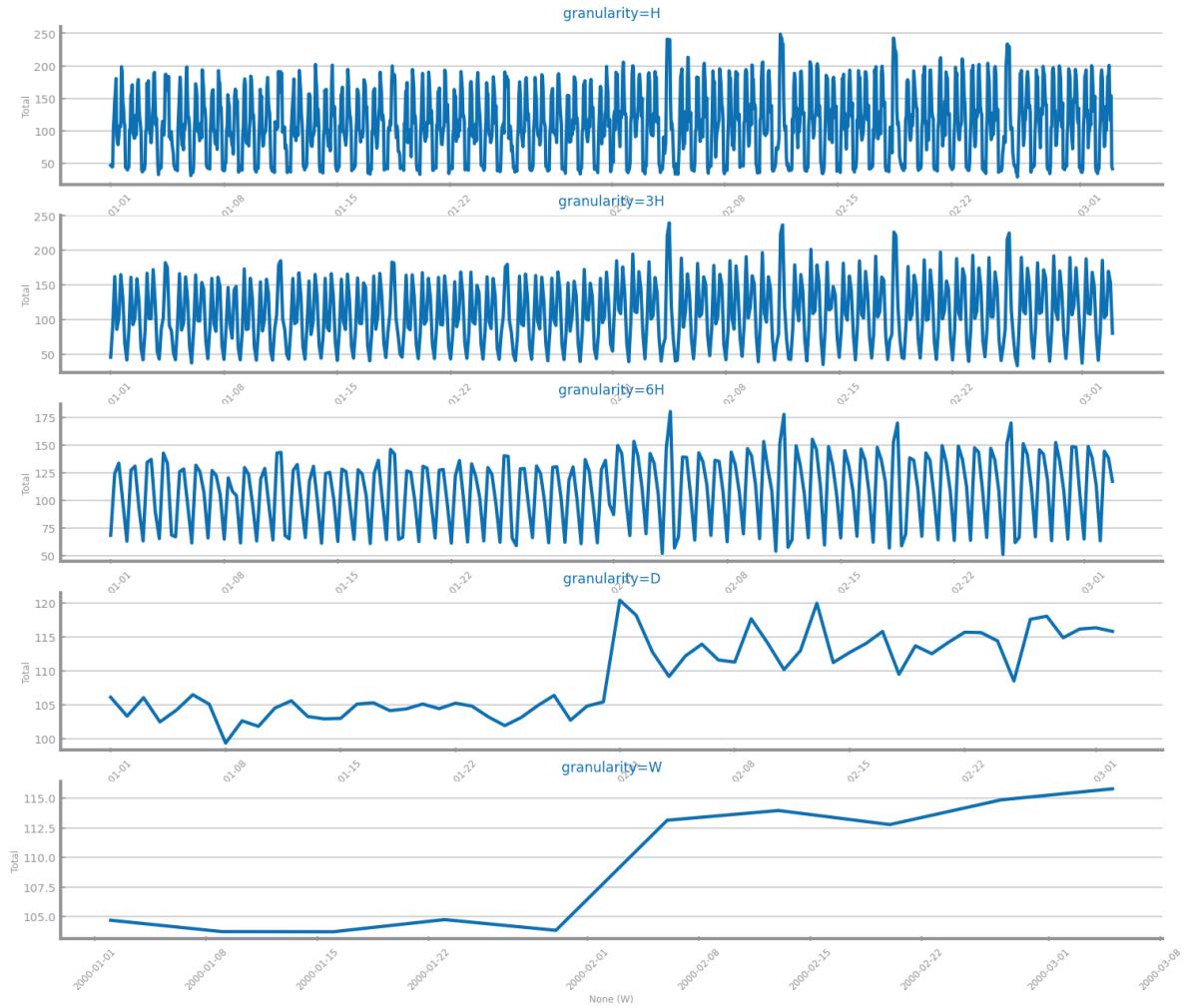


Figure 2: Traffic Time Series at three different granularities

## Data Distribution

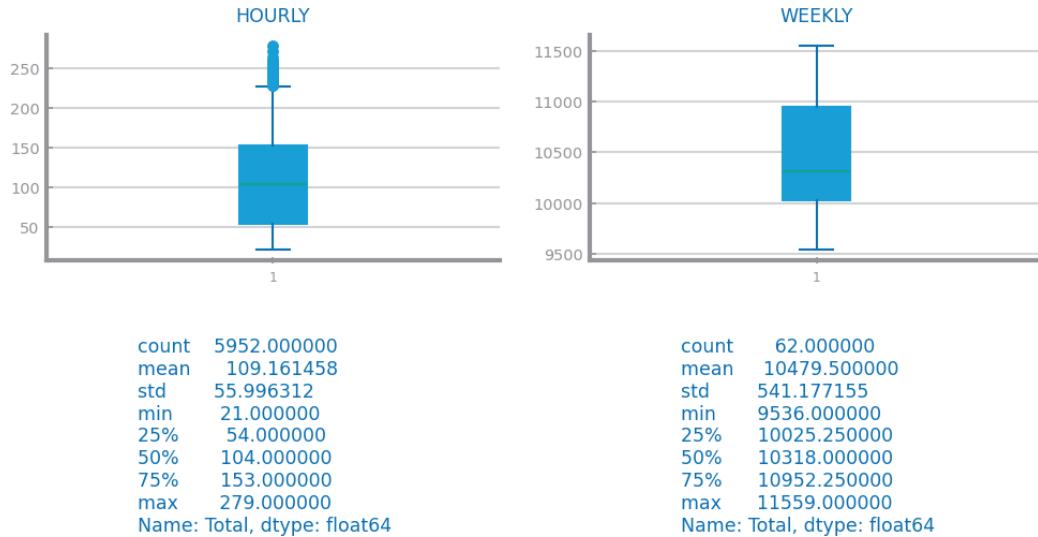


Figure 3: Boxplot(s) for Traffic Time Series

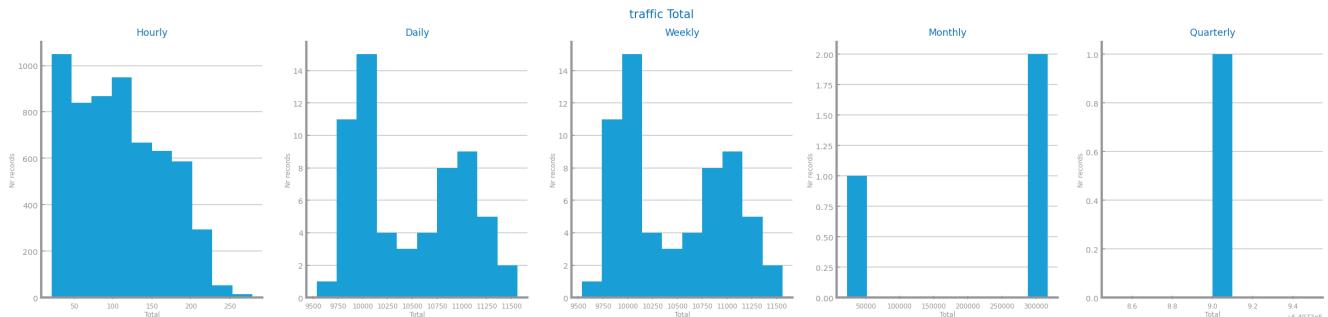


Figure 4: Histogram(s) for Traffic Time Series

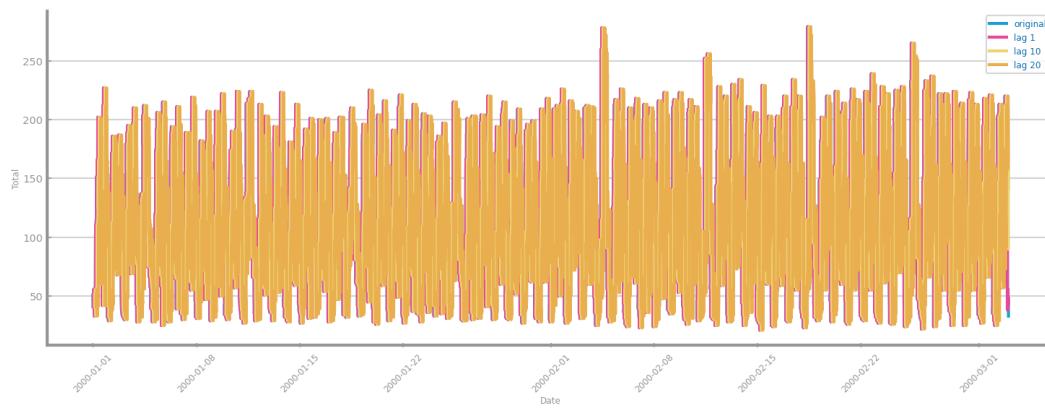


Figure 5: Autocorrelation lag-plots for original Traffic Time Series

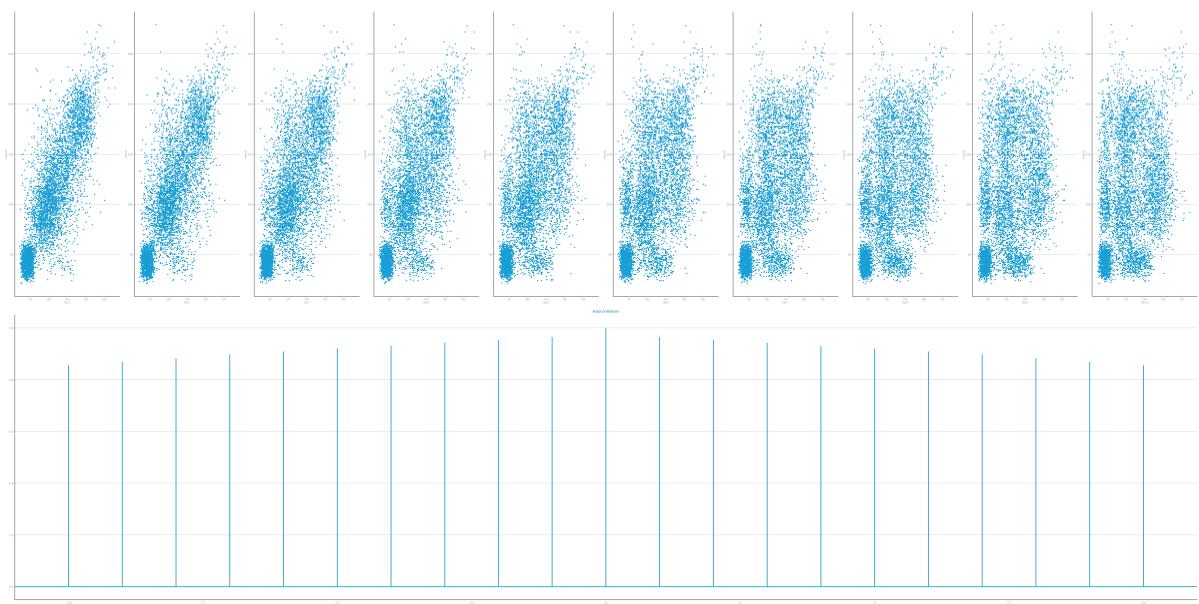


Figure 6: Autocorrelation correlogram for original Traffic Time Series

## Data Stationarity

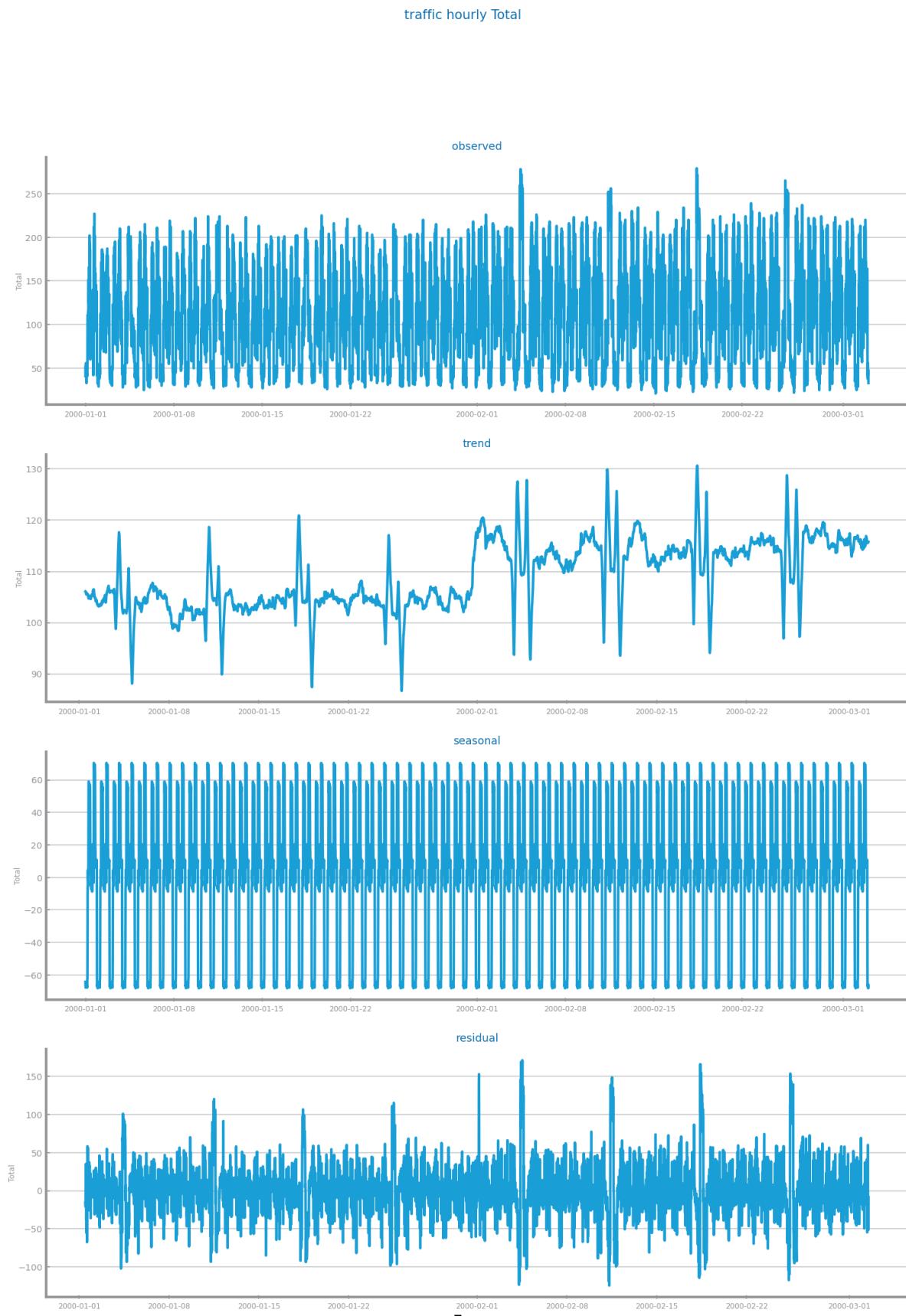


Figure 7: Components study for Traffic Time Series

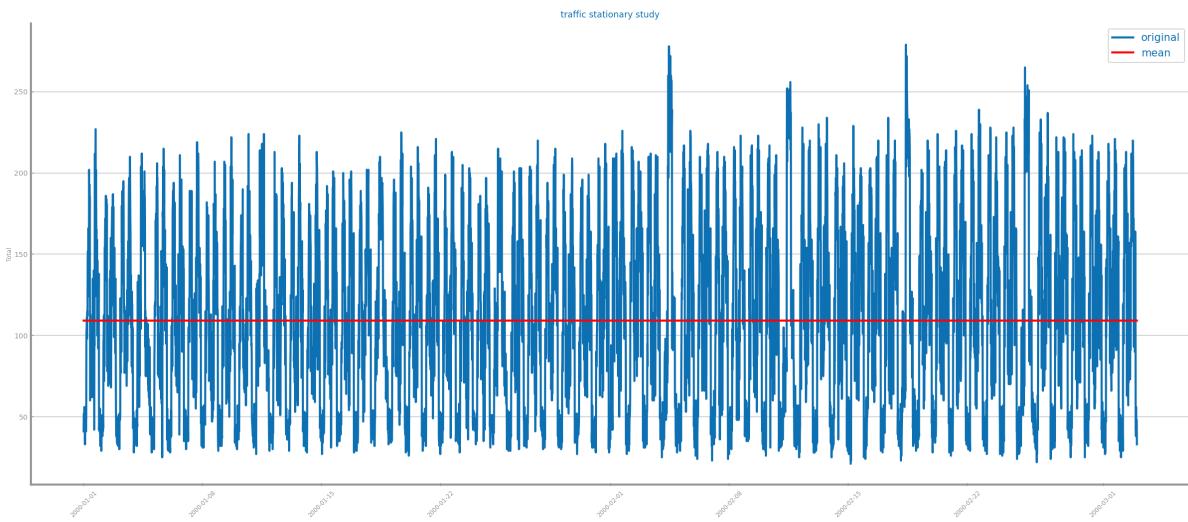


Figure 8: Stationarity study for Traffic Time Series

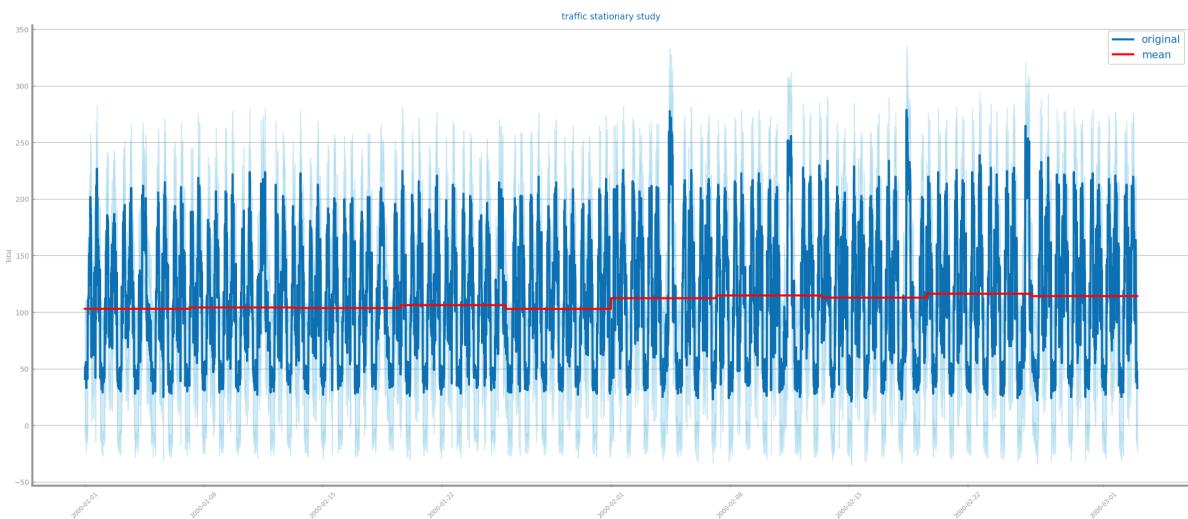


Figure 9: Stationarity study with bins for Traffic Time Series

# DATA TRANSFORMATION

## Aggregation



Figure 10: Forecasting predictions from Persistence Realist and Linear Regression on **Hourly Traffic Aggregation**

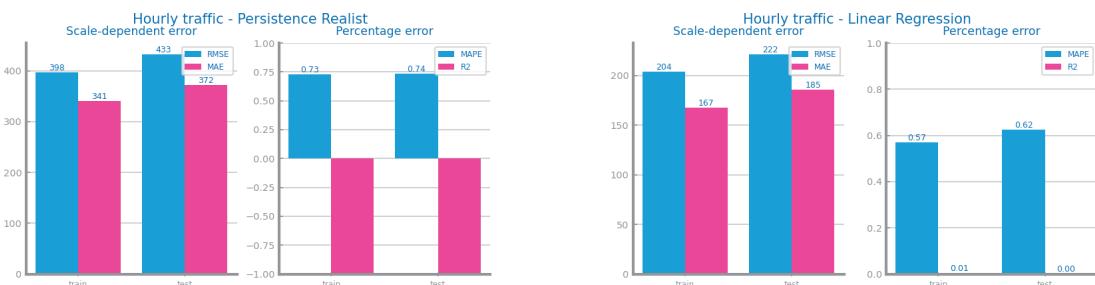


Figure 11: Performance of Persistence Realist and Linear Regression on **Hourly Traffic Aggregation**

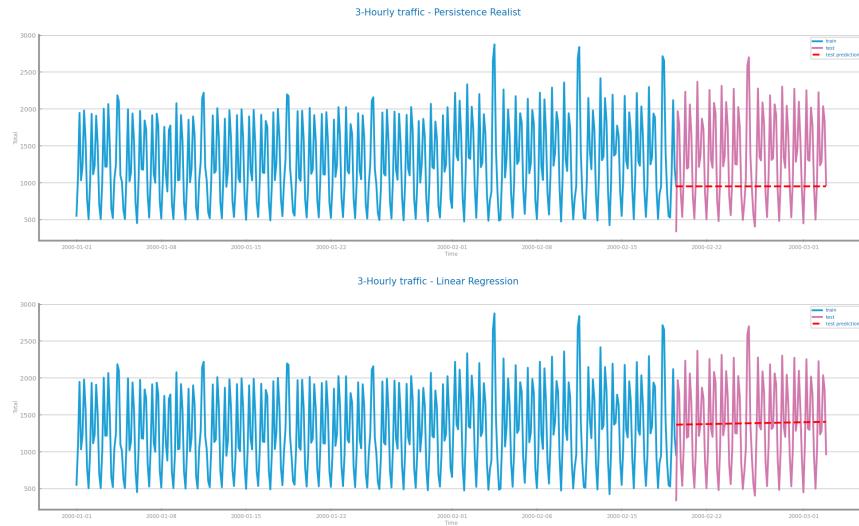


Figure 12: Forecasting predictions from Persistence Realist and Linear Regression on **3-Hourly Traffic Aggregation**

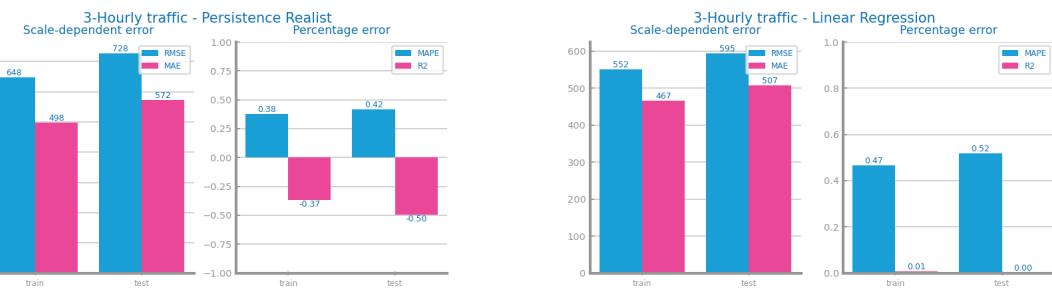


Figure 13: Performance of Persistence Realist and Linear Regression on **3-Hourly Traffic Aggregation**

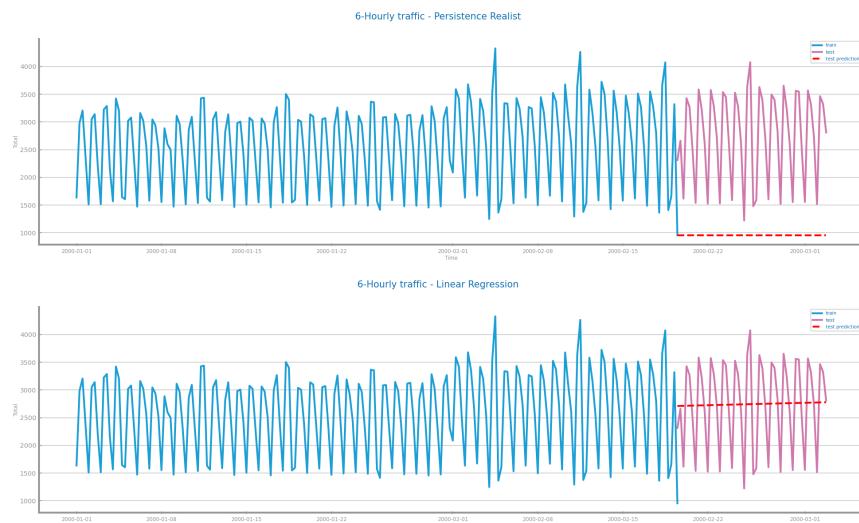


Figure 14: Forecasting predictions from Persistence Realist and Linear Regression on **6-Hourly Traffic Aggregation**

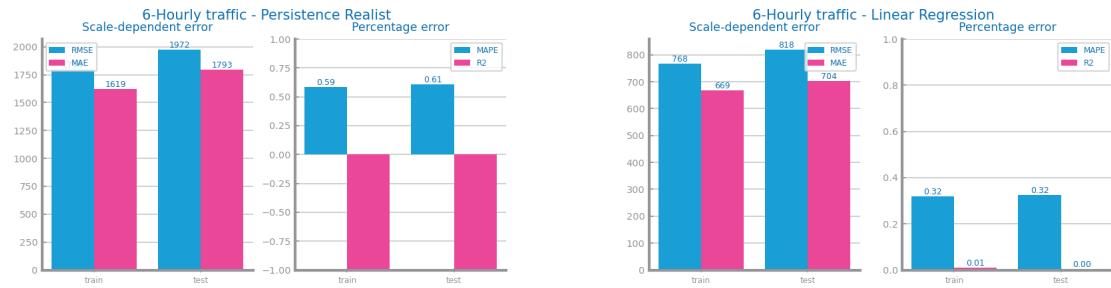


Figure 15: Performance of Persistence Realist and Linear Regression on **6-Hourly Traffic Aggregation**

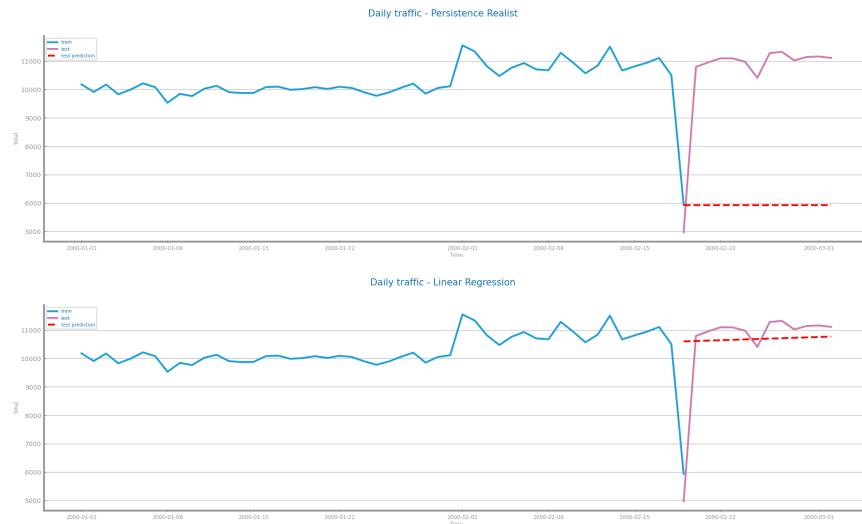


Figure 16: Forecasting predictions from Persistence Realist and Linear Regression on **Daily Traffic Aggregation**

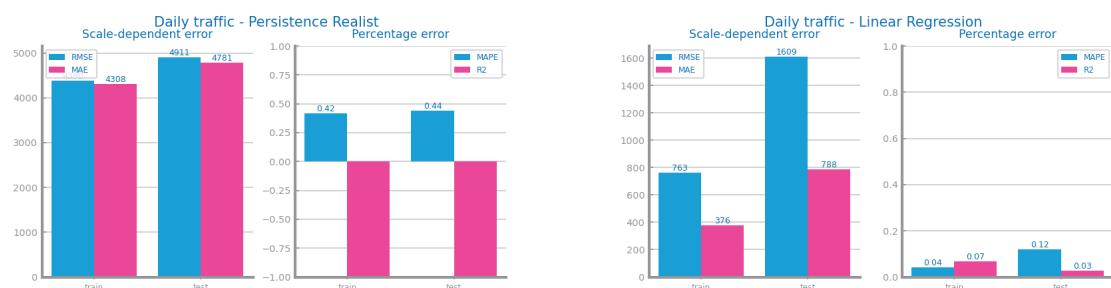


Figure 17: Performance of Persistence Realist and Linear Regression on **Daily Traffic Aggregation**

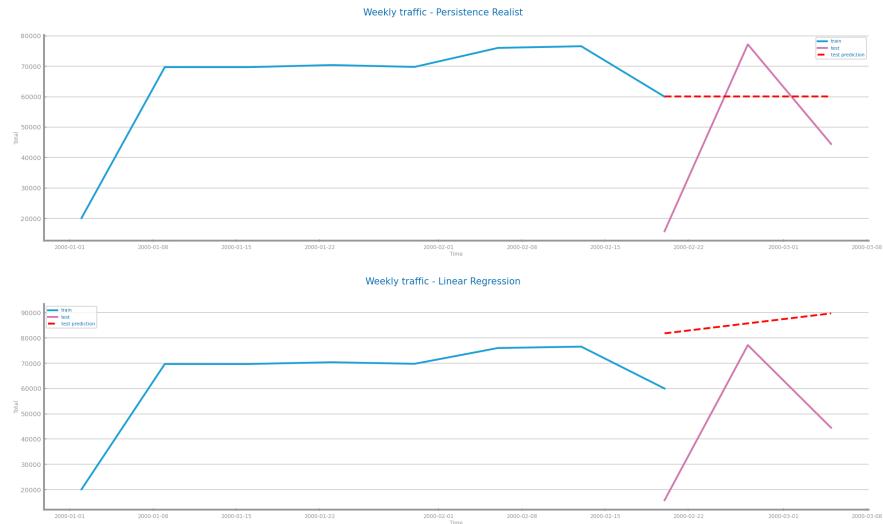


Figure 18: Forecasting predictions from Persistence Realist and Linear Regression on **Weekly Traffic Aggregation**

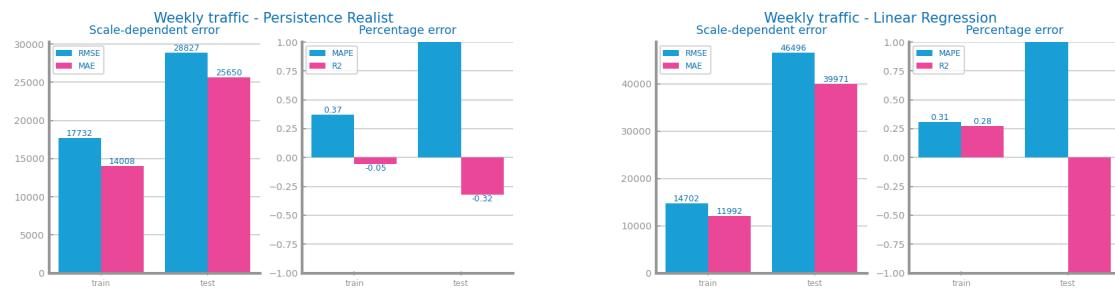


Figure 19: Performance of Persistence Realist and Linear Regression on **Weekly Traffic Aggregation**



Figure 20: Model comparison for different Traffic aggregation strategies

## *Differentiation*

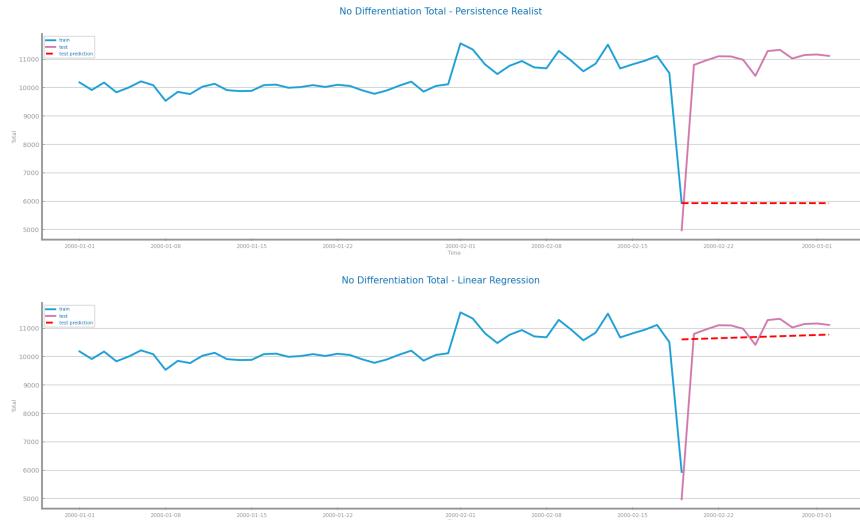


Figure 21: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic Baseline (No Differentiation)**



Figure 22: Performance of Persistence Realist and Linear Regression on **Traffic Baseline (No Differentiation)**

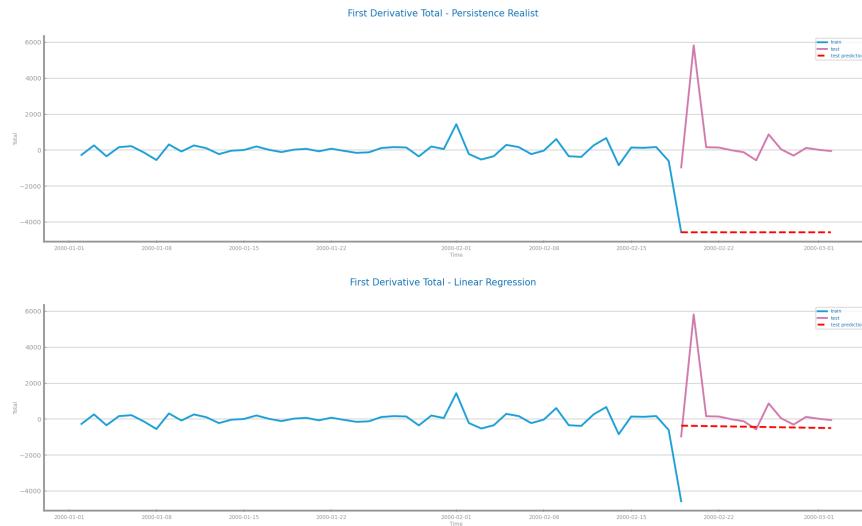


Figure 23: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic Differentiation (1st Order)**



Figure 24: Performance of Persistence Realist and Linear Regression on **Traffic Differentiation (1st Order)**



Figure 25: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic Differentiation (2nd Order)**

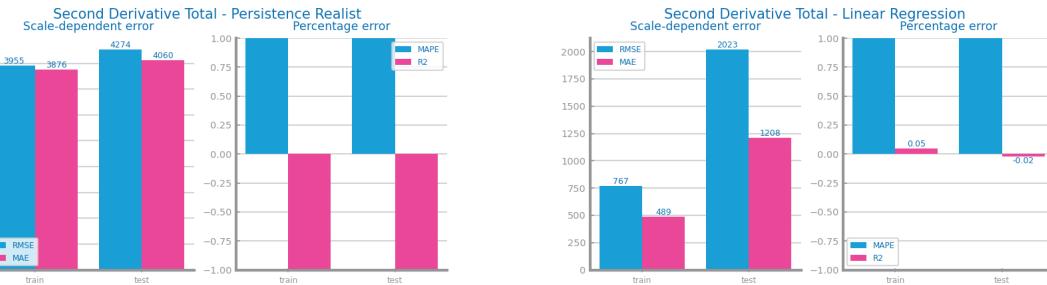


Figure 26: Performance of Persistence Realist and Linear Regression on **Traffic Differentiation (2nd Order)**



Figure 27: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic Differentiation with Lag 2**

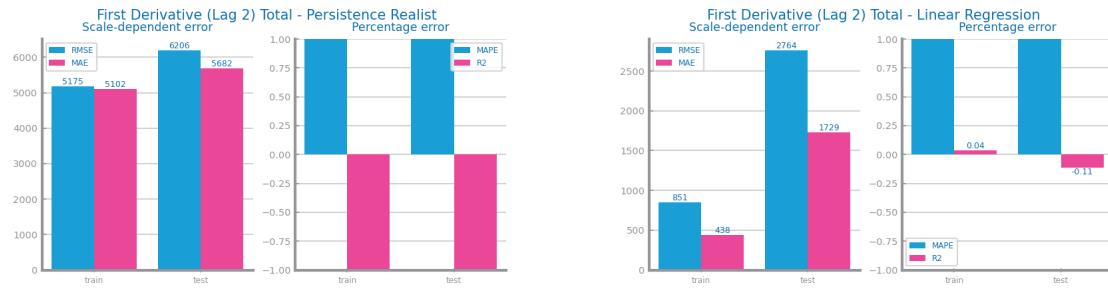


Figure 28: Performance of Persistence Realist and Linear Regression on **Traffic Differentiation with Lag 2**

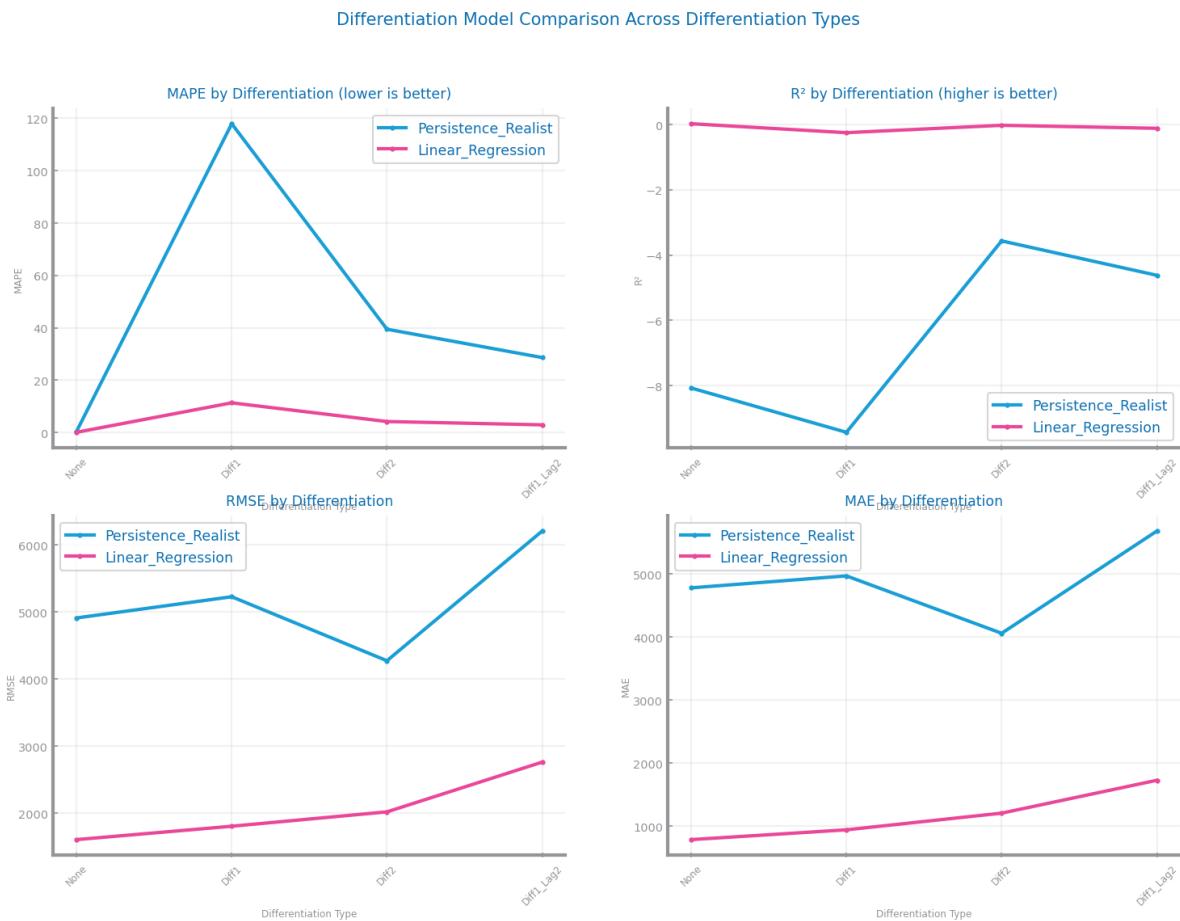


Figure 29: Model comparison for different Traffic differentiation strategies

## Smoothing



Figure 30: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with Smoothing Size 5**



Figure 31: Performance of Persistence Realist and Linear Regression on **Traffic with Smoothing Size 5**

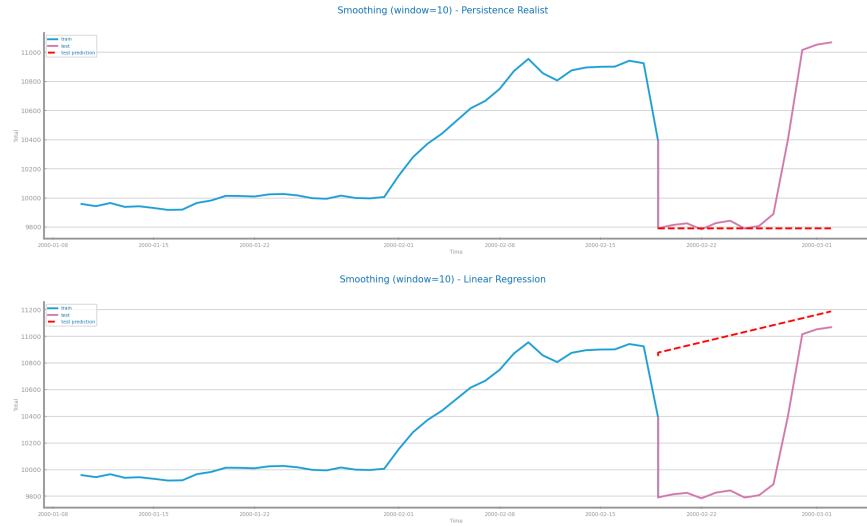


Figure 32: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with Smoothing Size 10**

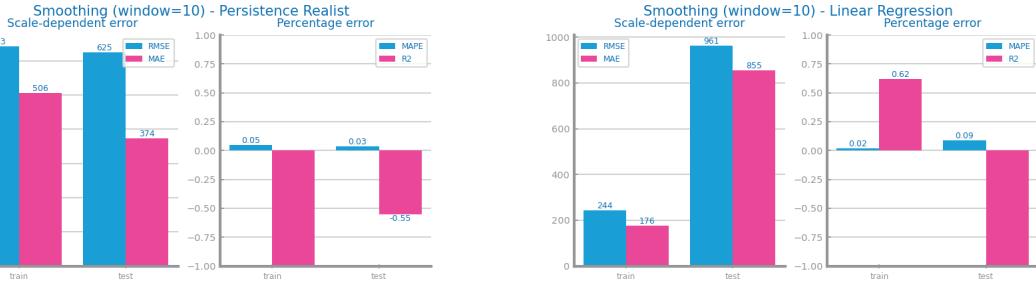


Figure 33: Performance of Persistence Realist and Linear Regression on **Traffic with Smoothing Size 10**



Figure 34: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with Smoothing Size 15**



Figure 35: Performance of Persistence Realist and Linear Regression on **Traffic with Smoothing Size 15**



Figure 36: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with Smoothing Size 25**

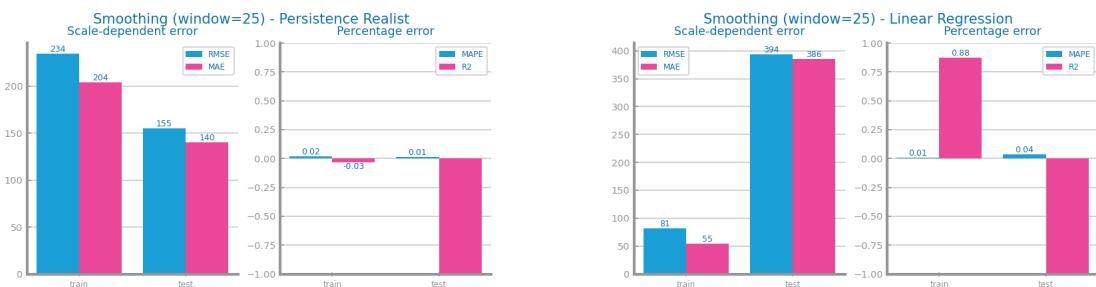


Figure 37: Performance of Persistence Realist and Linear Regression on **Traffic with Smoothing Size 25**

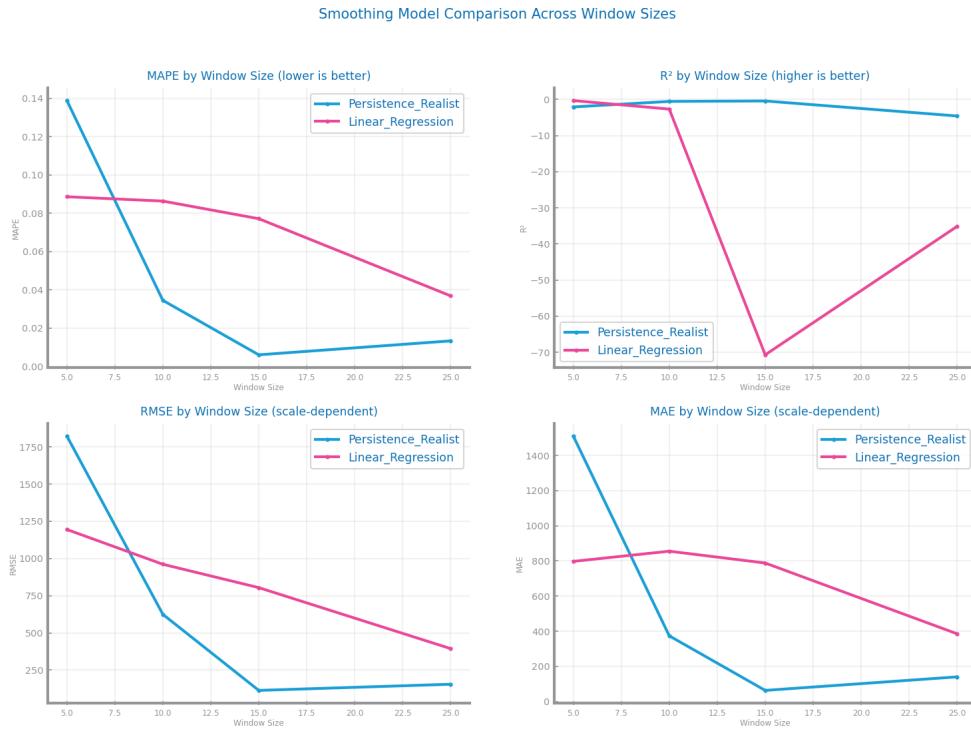


Figure 38: Model comparison for different Traffic smoothing strategies

## Scaling

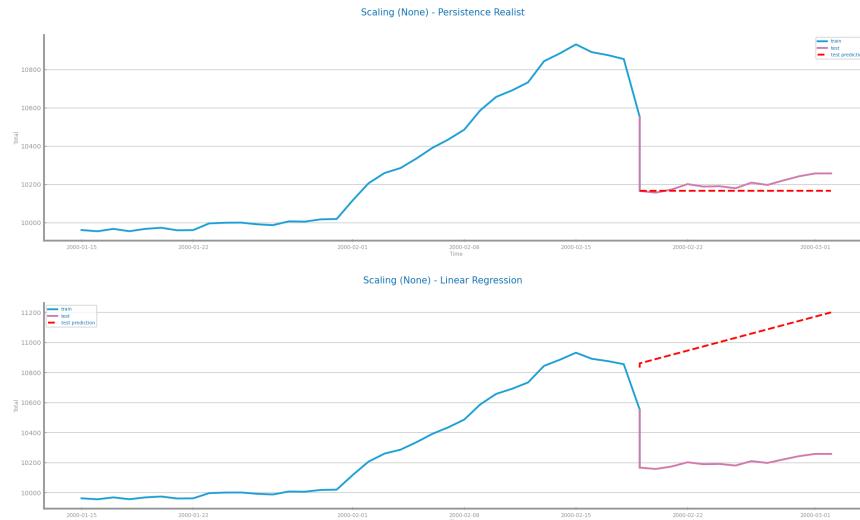


Figure 39: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with No Scaling**

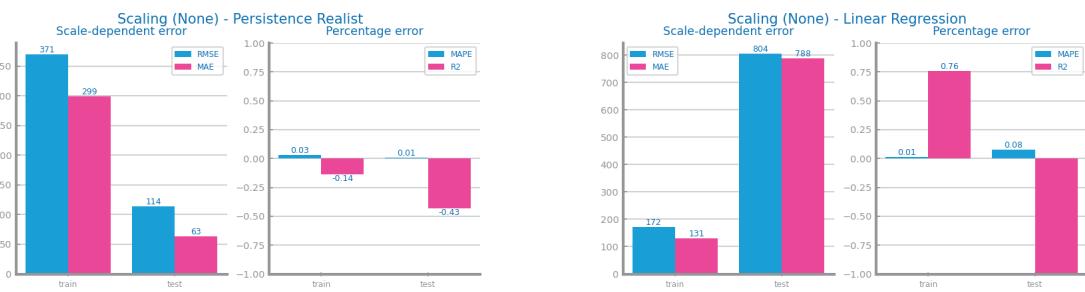


Figure 40: Performance of Persistence Realist and Linear Regression on **Traffic with No Scaling**



Figure 41: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with StandardScaler**

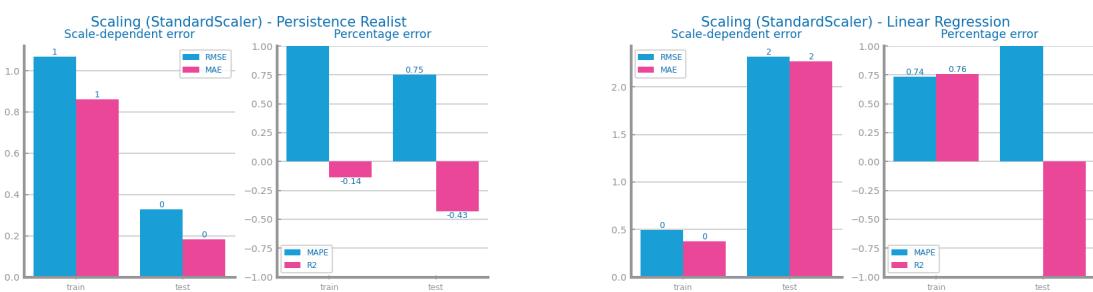


Figure 42: Performance of Persistence Realist and Linear Regression on **Traffic with StandardScaler**

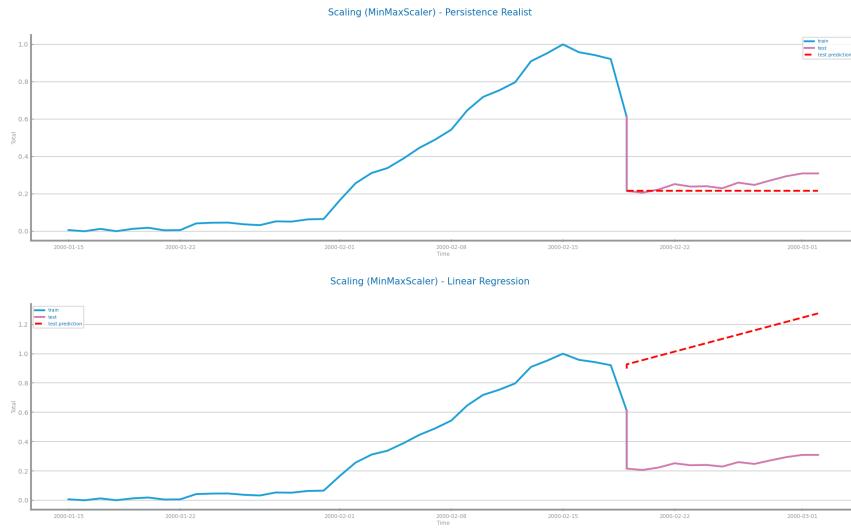


Figure 43: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with MinMaxScaler**



Figure 44: Performance of Persistence Realist and Linear Regression on **Traffic with MinMaxScaler**



Figure 45: Forecasting predictions from Persistence Realist and Linear Regression on **Traffic with RobustScaler**

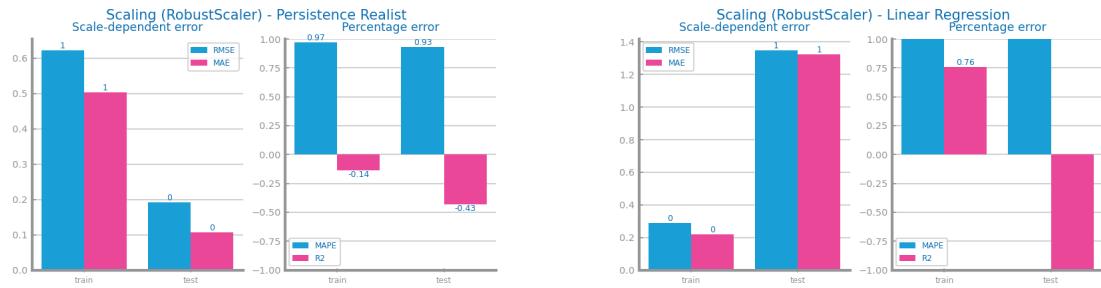


Figure 46: Performance of Persistence Realist and Linear Regression on **Traffic with RobustScaler**

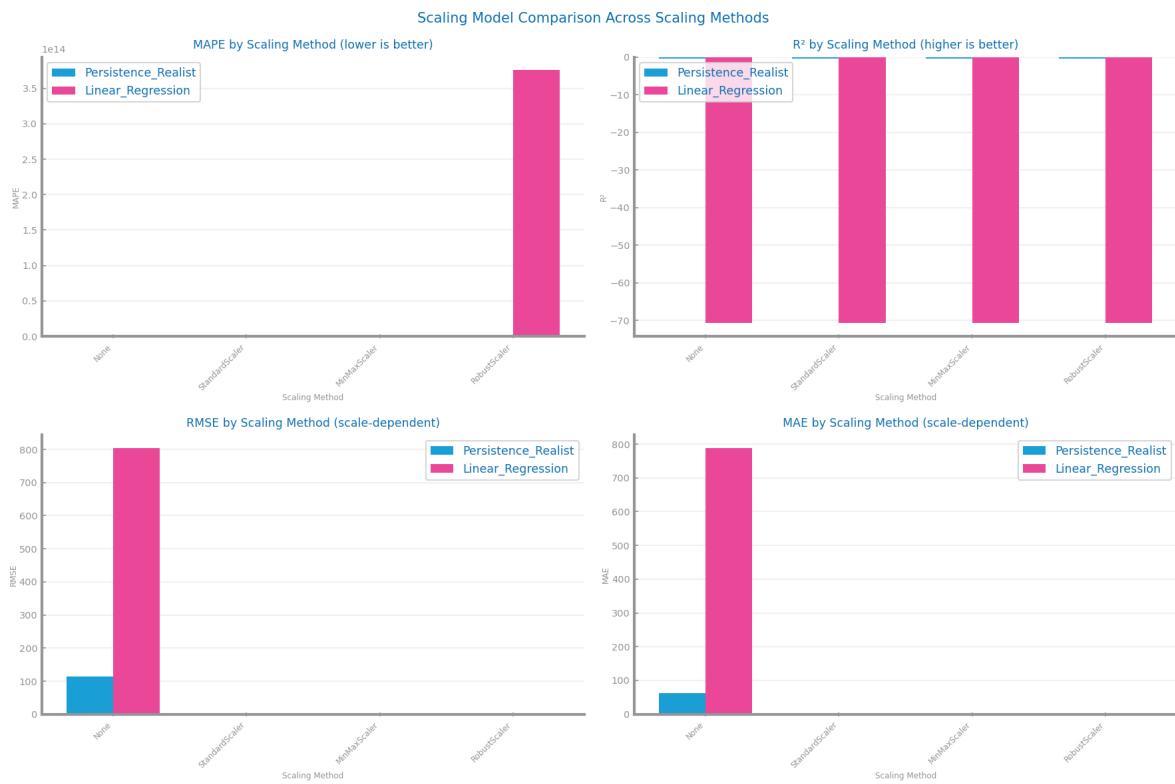


Figure 47: Model comparison for different Traffic scaling strategies

# Inflation Rate Time Series

## DATA PROFILING

### *Data Dimensionality and Granularity*



Figure 48: U.S. Inflation Rate dimensionality over time.

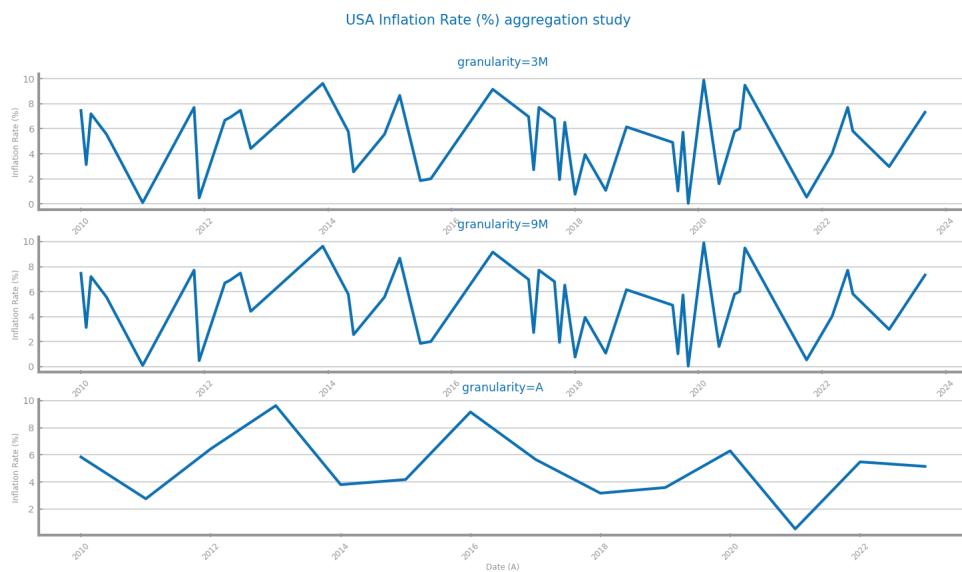


Figure 49: Inflation Rate Time Series at three different granularities

## Data Distribution

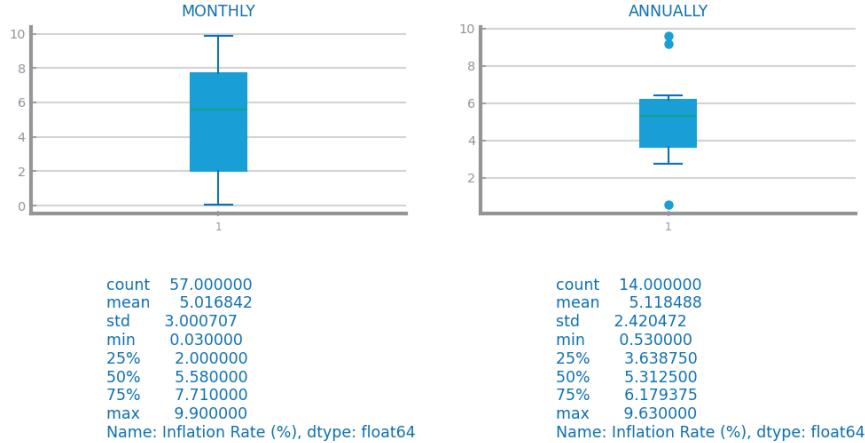


Figure 50: Boxplot(s) for Inflation Rate Time Series

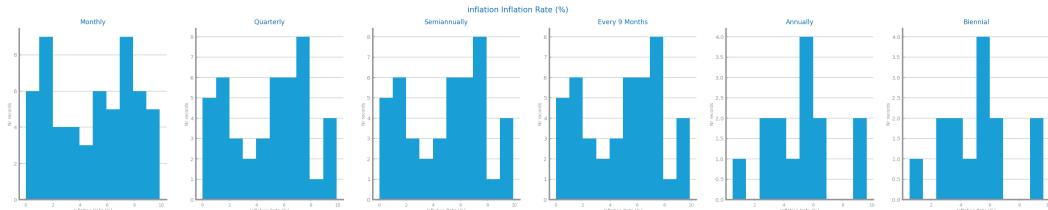


Figure 51: Histogram(s) for Inflation Rate Time Series

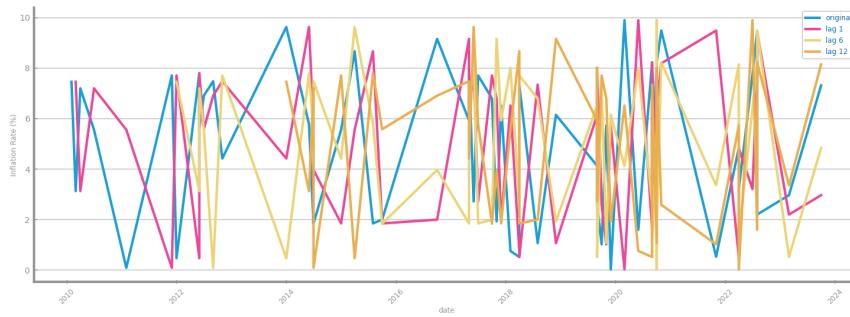


Figure 52: Autocorrelation lag-plots for original Inflation Rate Time Series

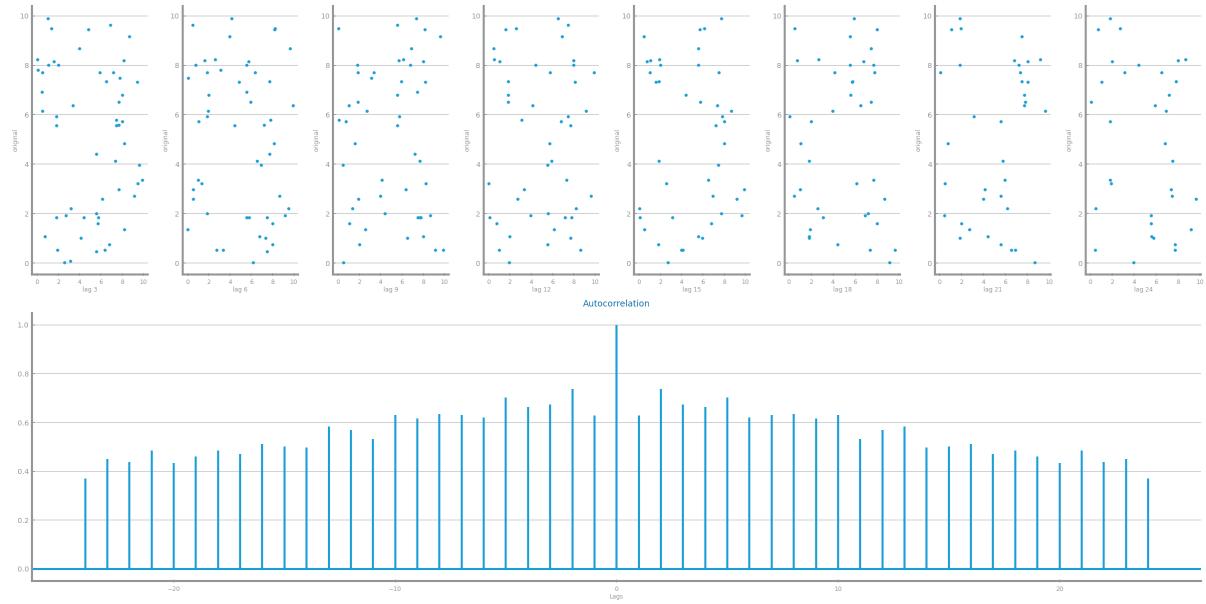


Figure 53: Autocorrelation correlogram for original Inflation Rate Time Series

## Data Stationarity

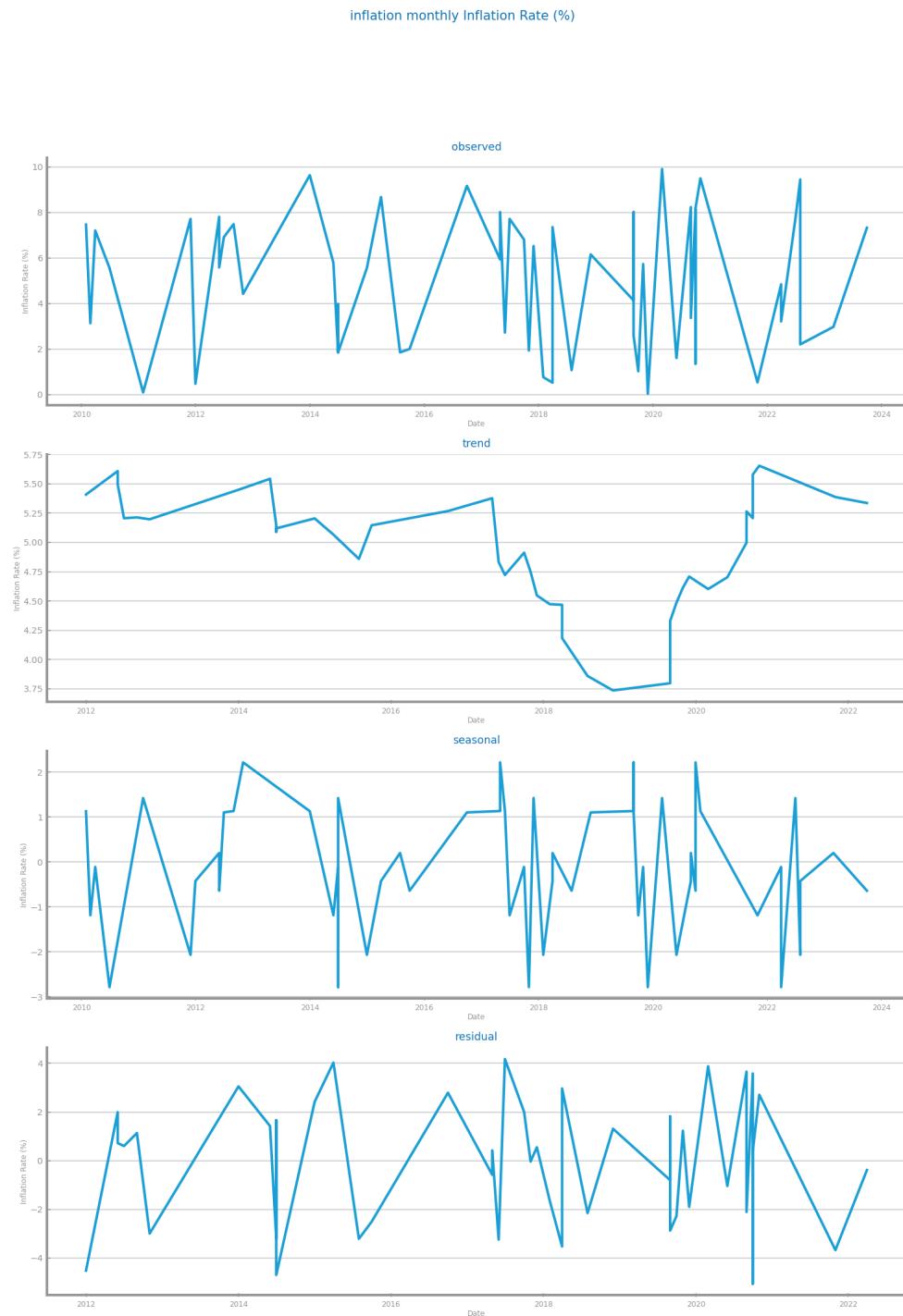


Figure 54: Components study for Inflation Rate Time Series

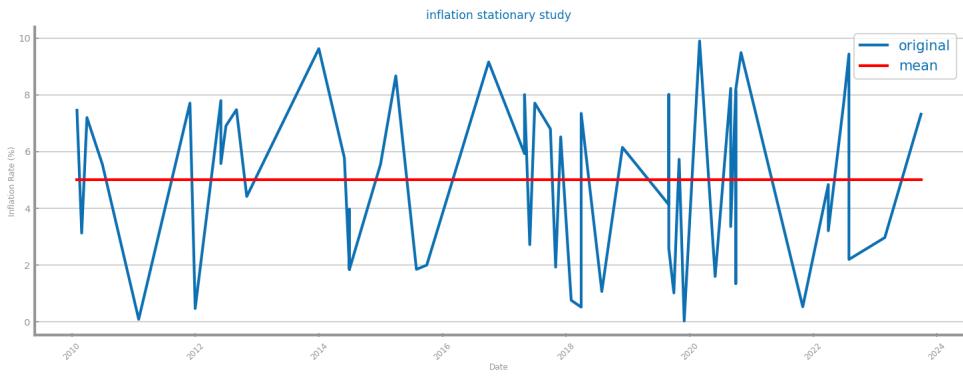


Figure 55: Stationarity study for Inflation Rate Time Series

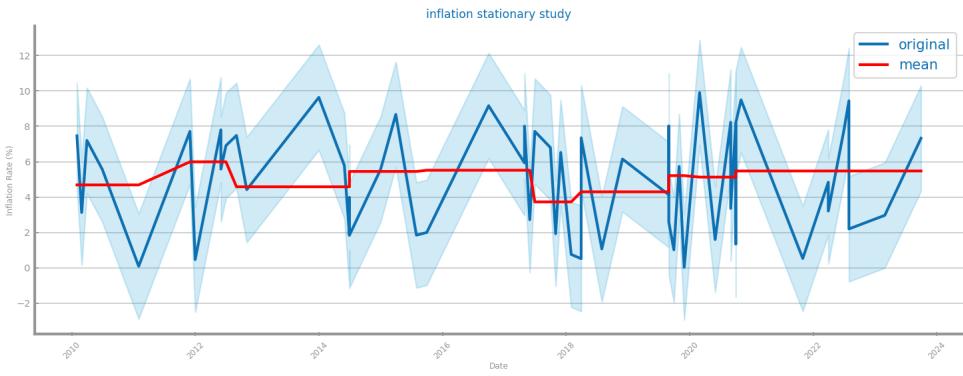


Figure 56: Stationarity study with bins for Inflation Rate Time Series

```

ADF Statistic: -10.410
p-value: 0.000
Critical Values:
    1%: -3.553
    5%: -2.915
    10%: -2.595
The series is stationary

```

Figure 57: Augmented Dickey-Fuller test results for Inflation Rate Time Series

# DATA TRANSFORMATION

## *Aggregation*

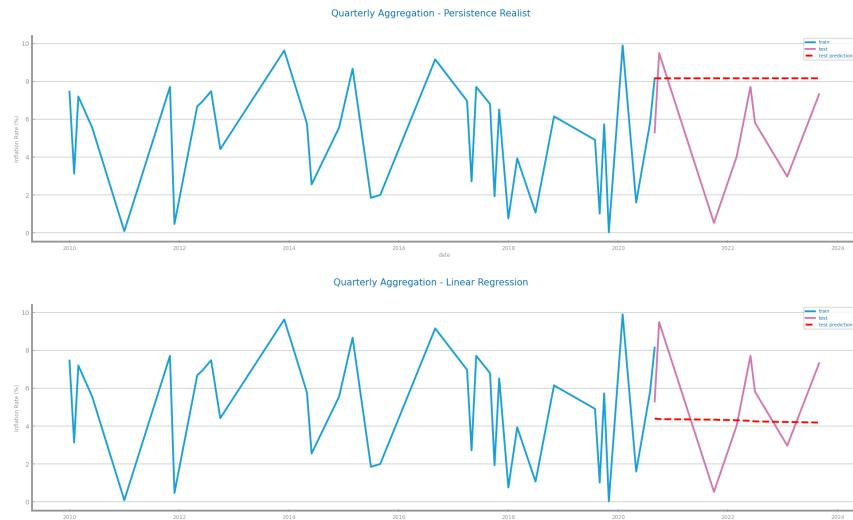


Figure 58: Forecasting predictions from Persistence Realist and Linear Regression on **Quarterly Inflation Aggregation**

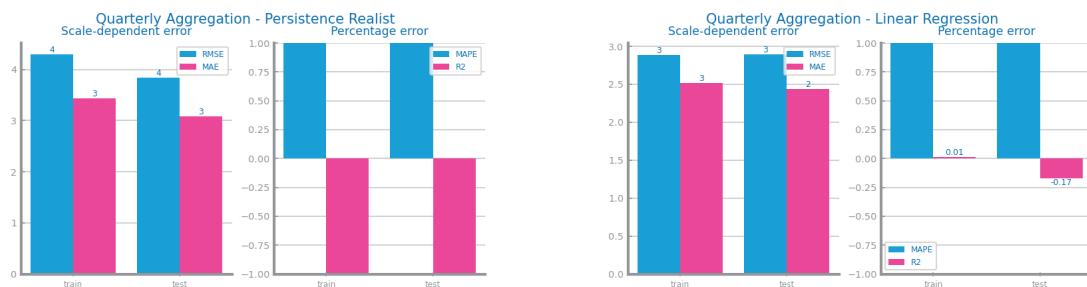


Figure 59: Performance of Persistence Realist and Linear Regression on **Quarterly Inflation Aggregation**

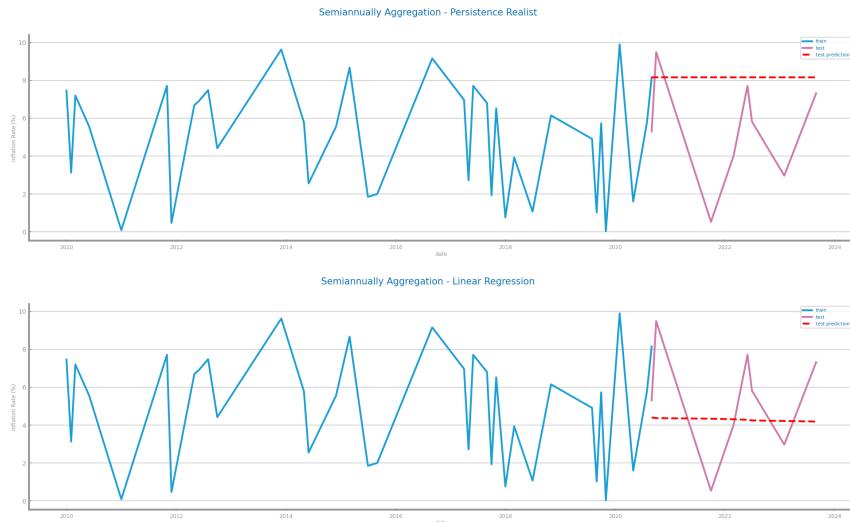


Figure 60: Forecasting predictions from Persistence Realist and Linear Regression on **Semi-Annual Inflation Aggregation**

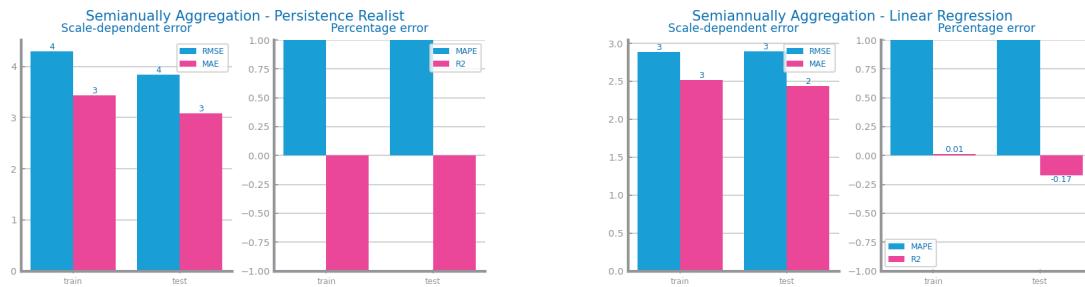


Figure 61: Performance of Persistence Realist and Linear Regression on **Semi-Annual Inflation Aggregation**

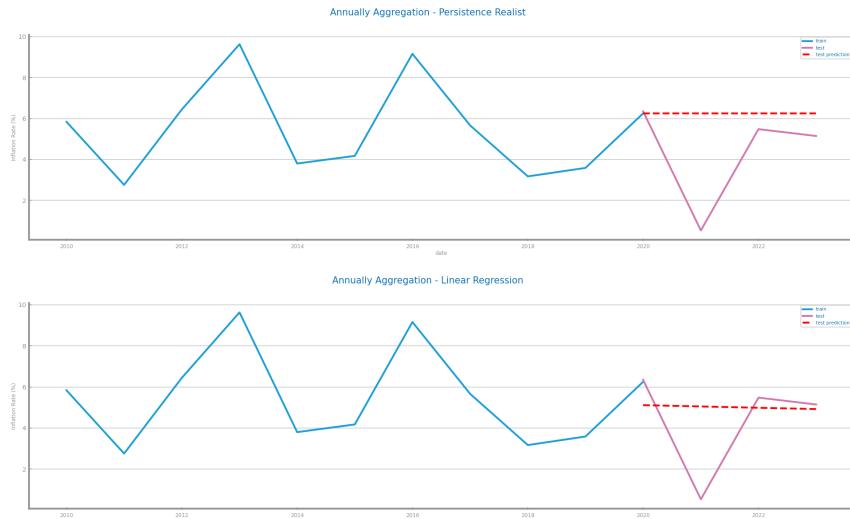


Figure 62: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Aggregation**

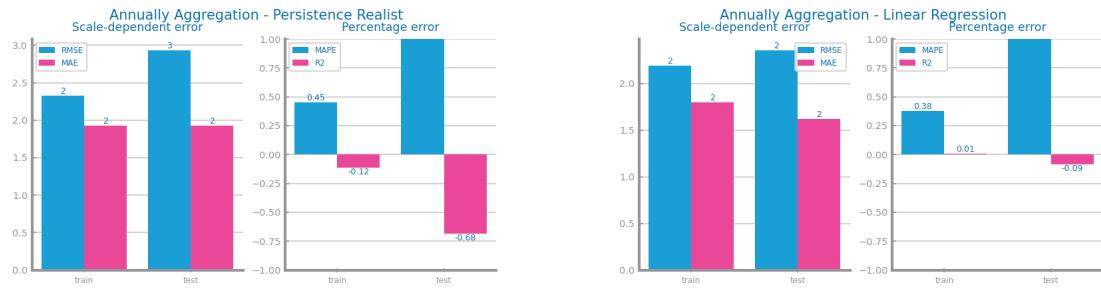


Figure 63: Performance of Persistence Realist and Linear Regression on **Annual Inflation Aggregation**

For all aggregation techniques, the Linear Regression model yields a lower MAE and RMSE than the Persistent Realist Model for the test set. Among the Linear Regression models, the Annual aggregation yields the lowest MAE for the test set (1.6).

### Differentiation

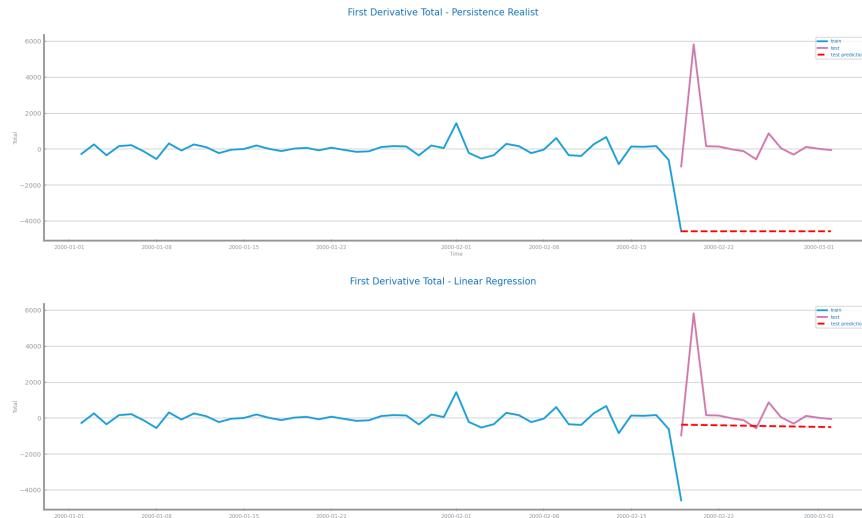


Figure 64: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Differentiation (1st Order)**



Figure 65: Performance of Persistence Realist and Linear Regression on **Annual Inflation Differentiation (1st Order)**



Figure 66: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Differentiation (2nd Order)**

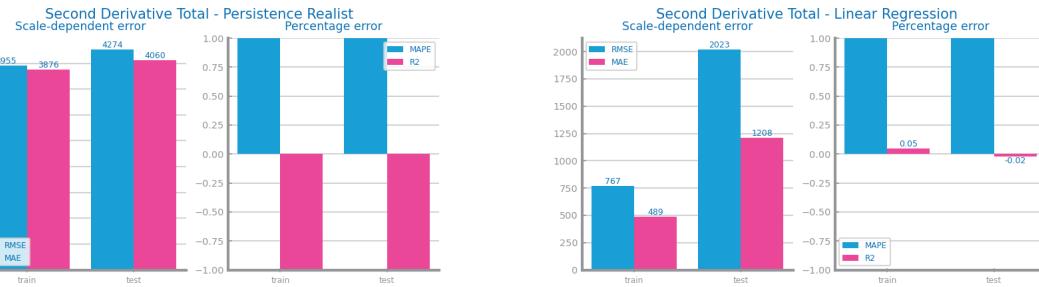


Figure 67: Performance of Persistence Realist and Linear Regression on **Annual Inflation Differentiation (2nd Order)**



Figure 68: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Differentiation with Lag 2**

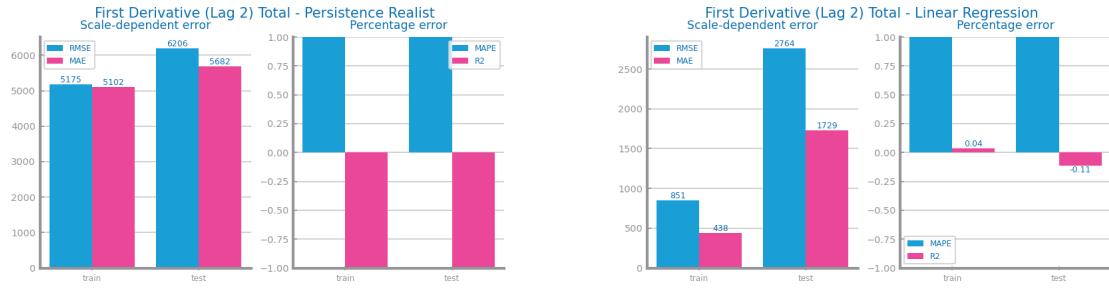


Figure 69: Performance of Persistence Realist and Linear Regression on **Annual Inflation Differentiation with Lag 2**

for diff\_1 and diff\_2, the Linear Regression Model yields a lower MAE and RMSE for the test set than the Persistent Realist models. For the diff\_lag2 treatment, the Persistent Realist has a lower MAE (but higher RMSE) than the Linear Regression Model. But overall, the best model, a Linear Regression Model trained on the diff\_1 treatment, is worse than the best model without differential treatment. So, we do not differentiate.

### Smoothing

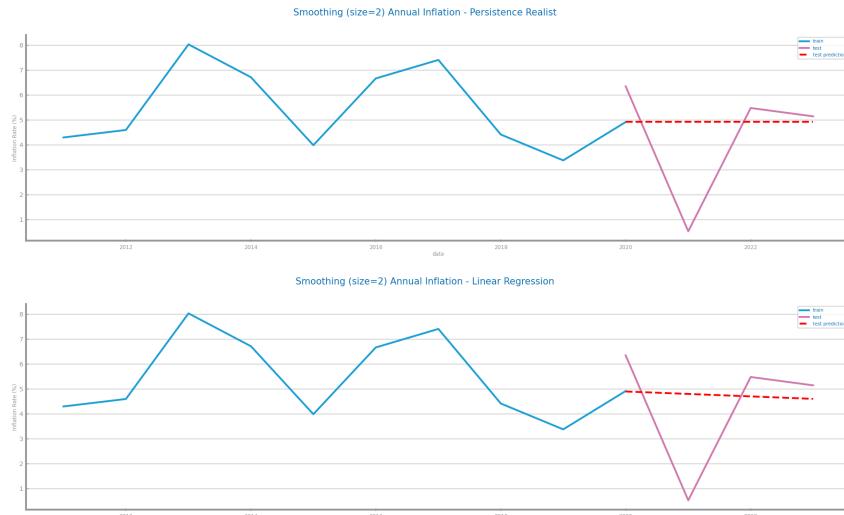


Figure 70: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=2)**

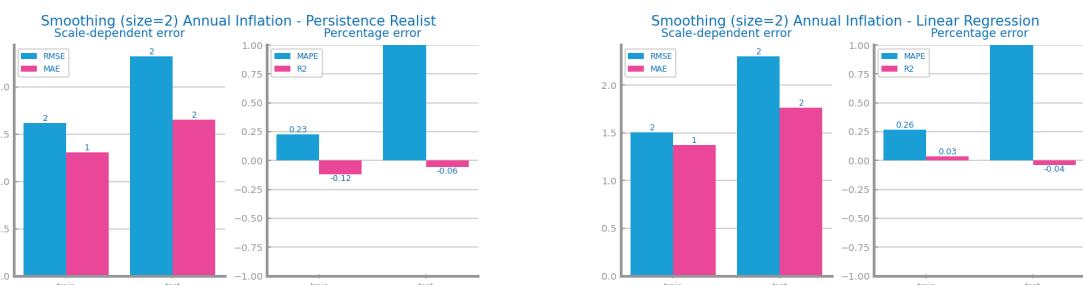


Figure 71: Performance of Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=2)**



Figure 72: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=3)**

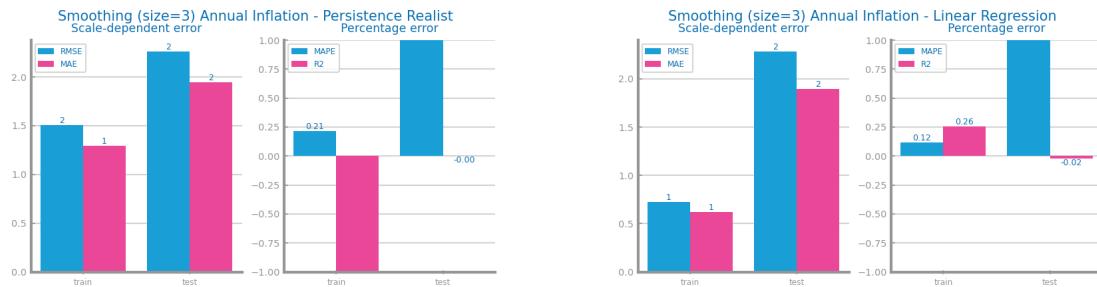


Figure 73: Performance of Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=3)**

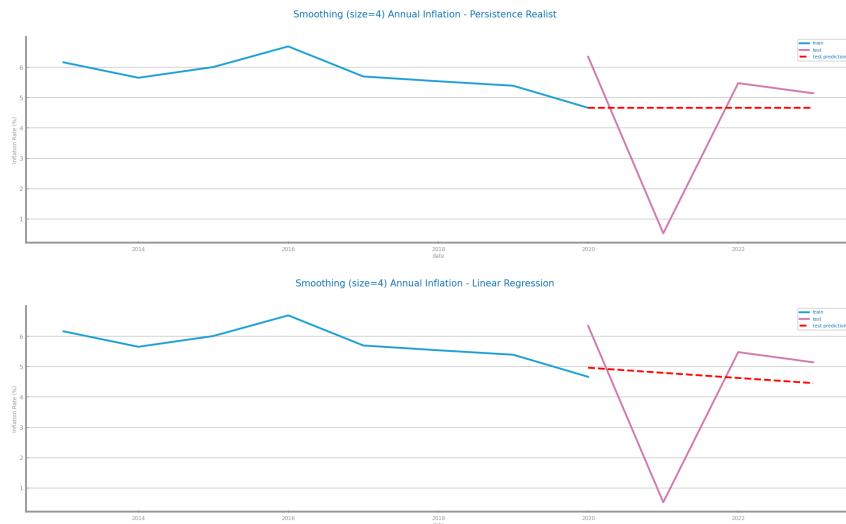


Figure 74: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=4)**

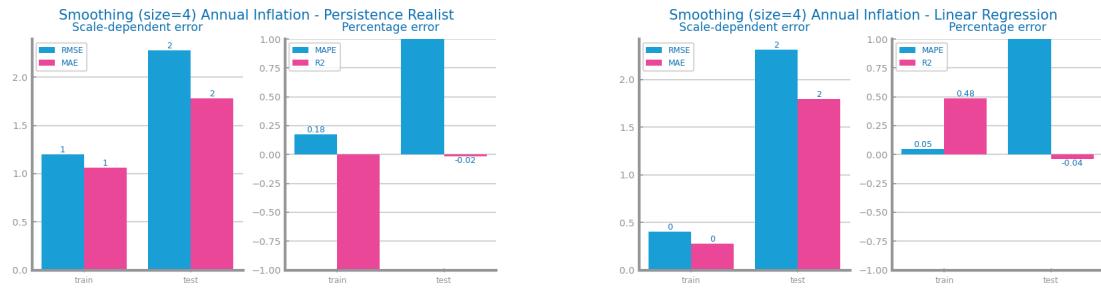


Figure 75: Performance of Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=4)**

Choosing the smoothing with `windowSize=2` since the Persistence Scored better than the best after aggregation and differentiation.

## Scaling

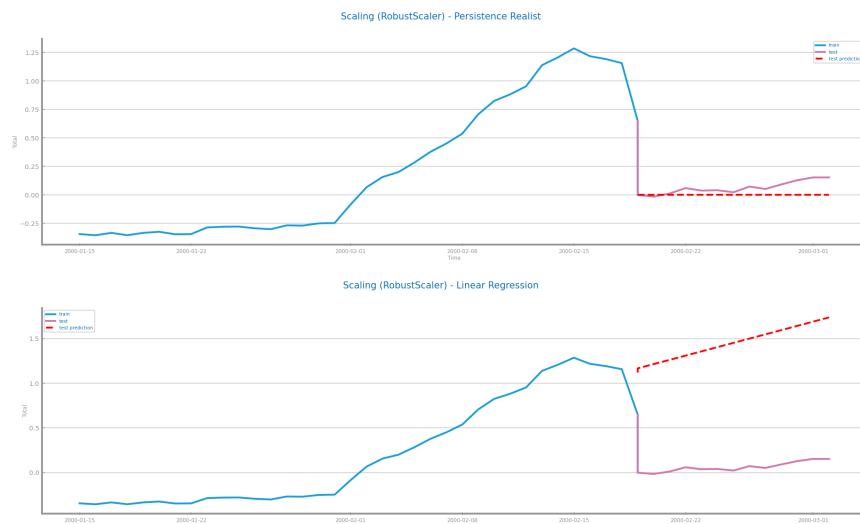


Figure 76: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Scaling - RobustScaler**



Figure 77: Performance of Persistence Realist and Linear Regression on **Annual Inflation Scaling - RobustScaler**

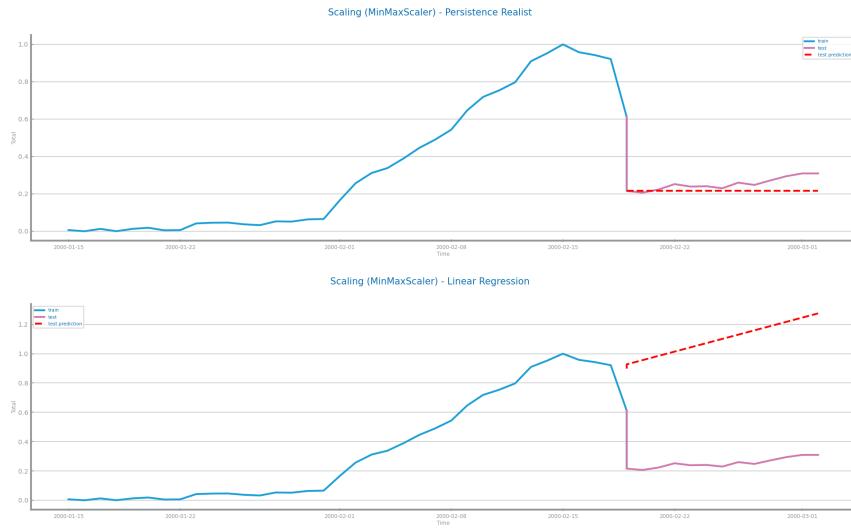


Figure 78: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Scaling - MinMaxScaler**



Figure 79: Performance of Persistence Realist and Linear Regression on **Annual Inflation Scaling - MinMaxScaler**



Figure 80: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Scaling - StandardScaler**



Figure 81: Performance of Persistence Realist and Linear Regression on **Annual Inflation Scaling - StandardScaler**

The best results are achieved with MinMaxScaler combined with the Persistence Realist model, due to its lowest Test MAE (0.35) and significantly better Test MAPE (0.61).