

# Data Science Project

<b>Team nr:</b> 16	<b>Student 1:</b> Antero Morgado <b>IST nr:</b> 1119213
	<b>Student 2:</b> David Ferreira <b>IST nr:</b> 1107077
	<b>Student 3:</b> José Fernandes <b>IST nr:</b> 1103727
	<b>Student 4:</b> Olha Buts <b>IST nr:</b> 1116276

# TIME SERIES ANALYSIS

## 1 DATA PROFILING

### *Data Dimensionality and Granularity*

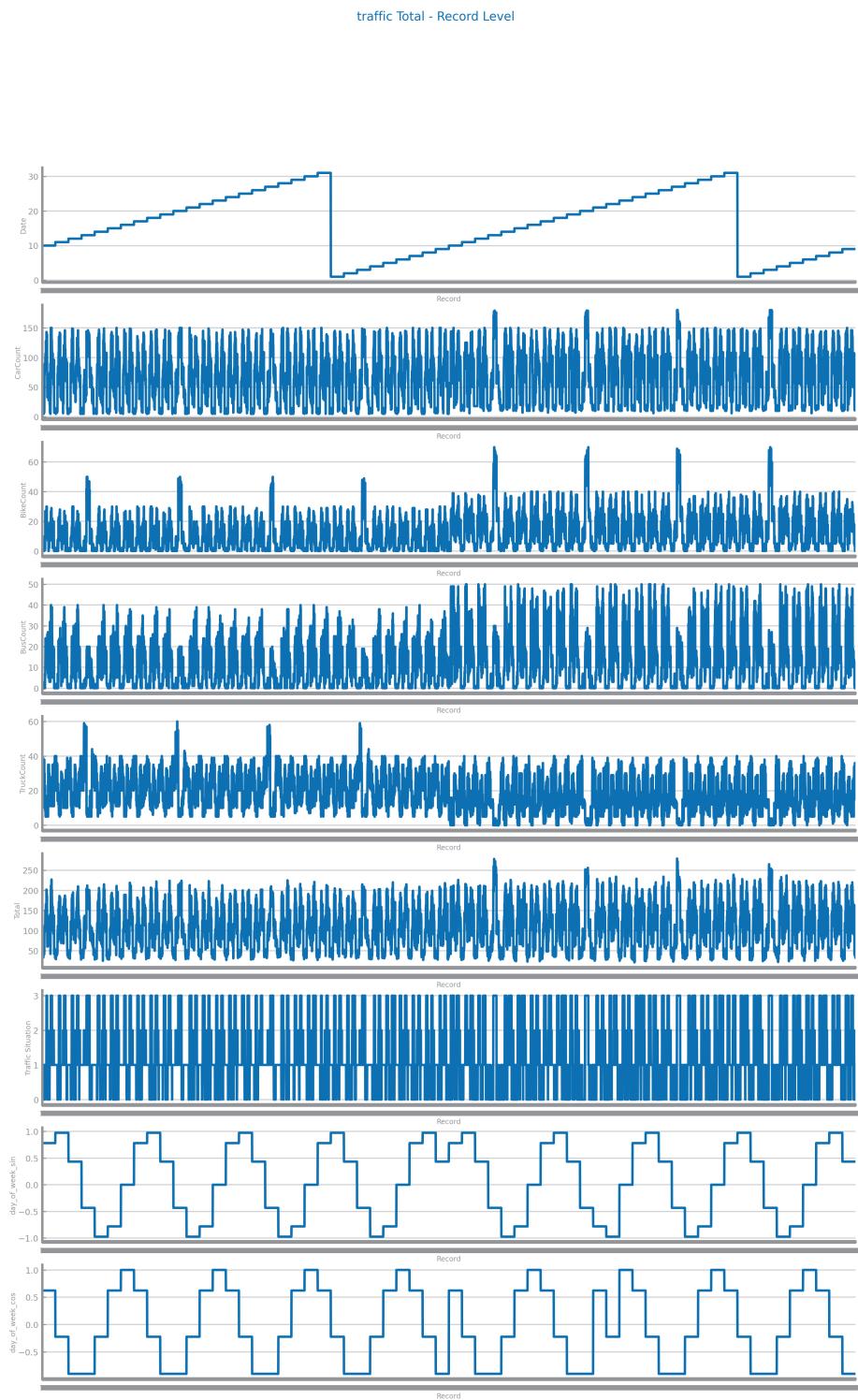


Figure 1: Time series 1 at the most granular detail

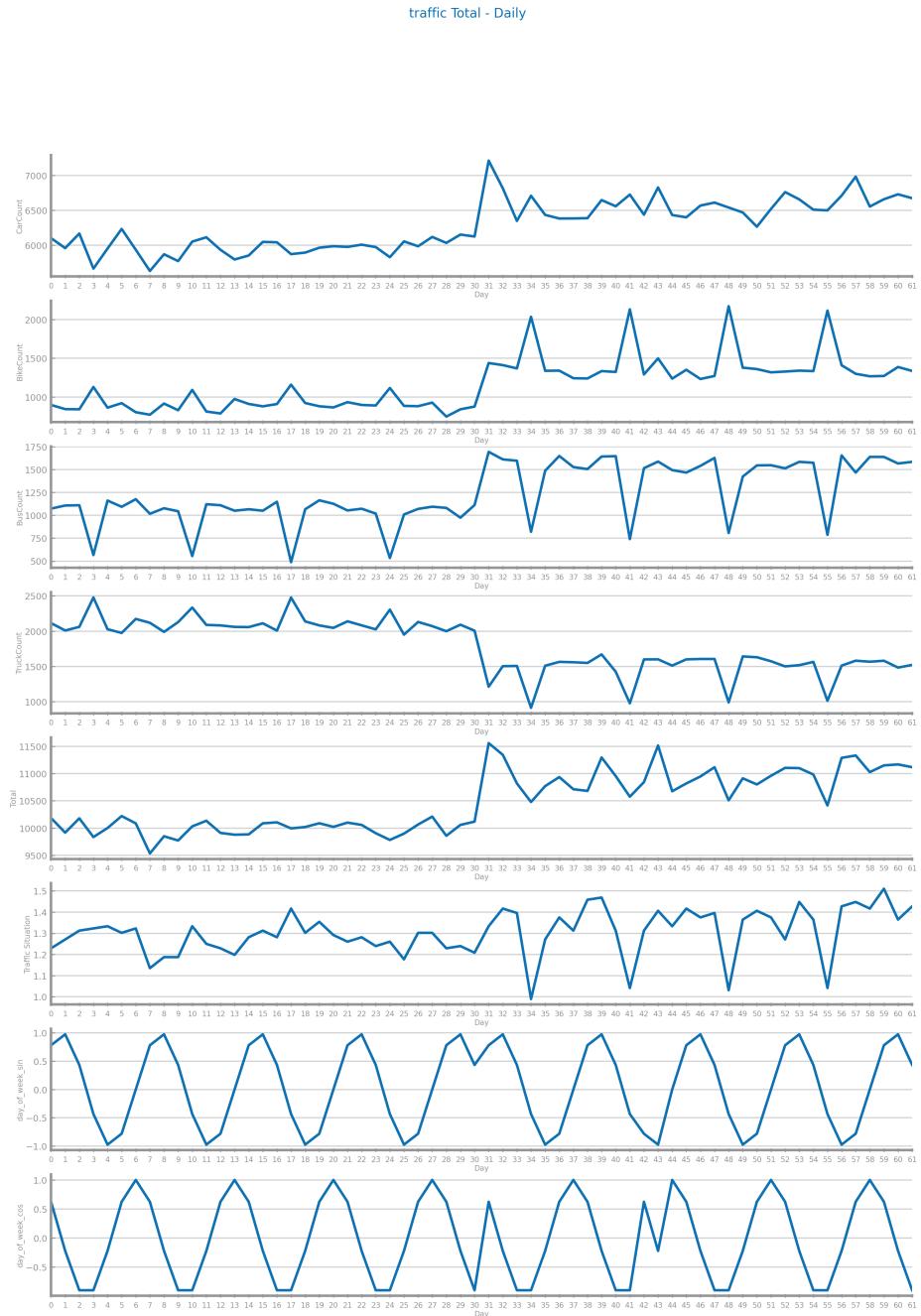


Figure 2: Time series 1 at the second chosen granularity

traffic Total - Weekly

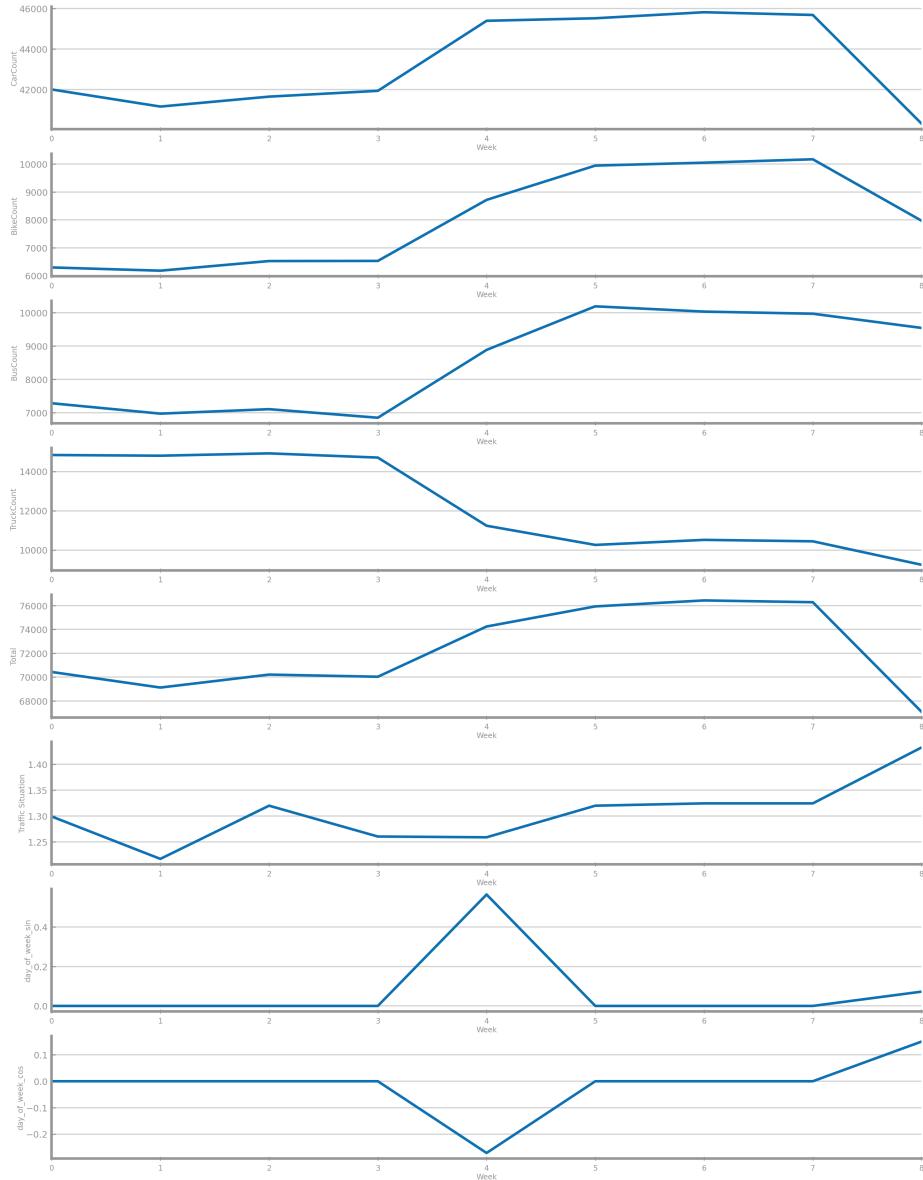


Figure 3: Time series 1 at the third chosen granularity

## Data Distribution

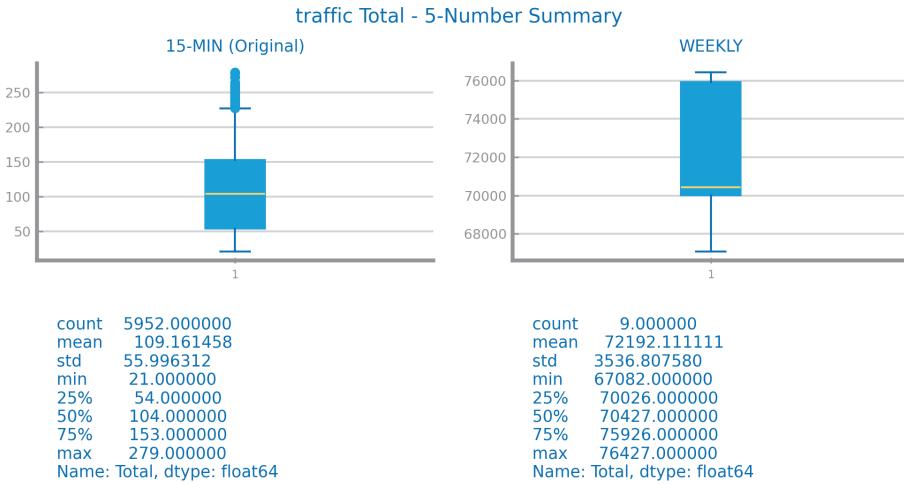


Figure 4: Boxplot(s) for time series 1

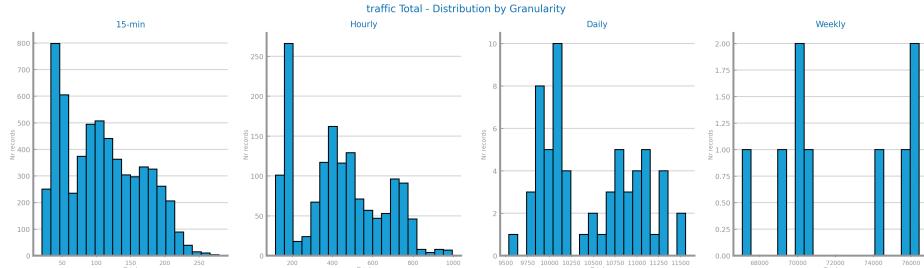


Figure 5: Histogram(s) for time series 1

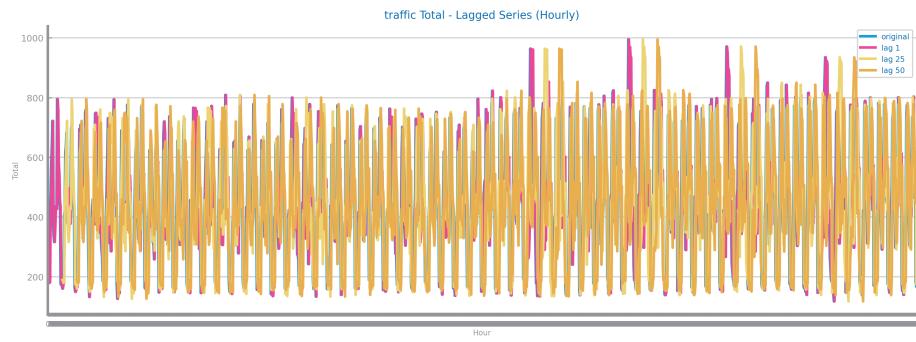


Figure 6: Autocorrelation lag-plots for original time series 1

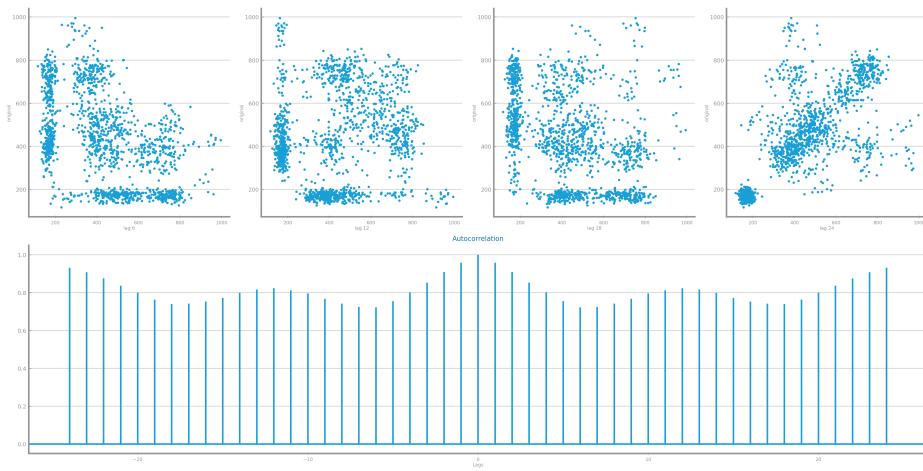


Figure 7: Autocorrelation correlogram for original time series 1

## Data Stationarity

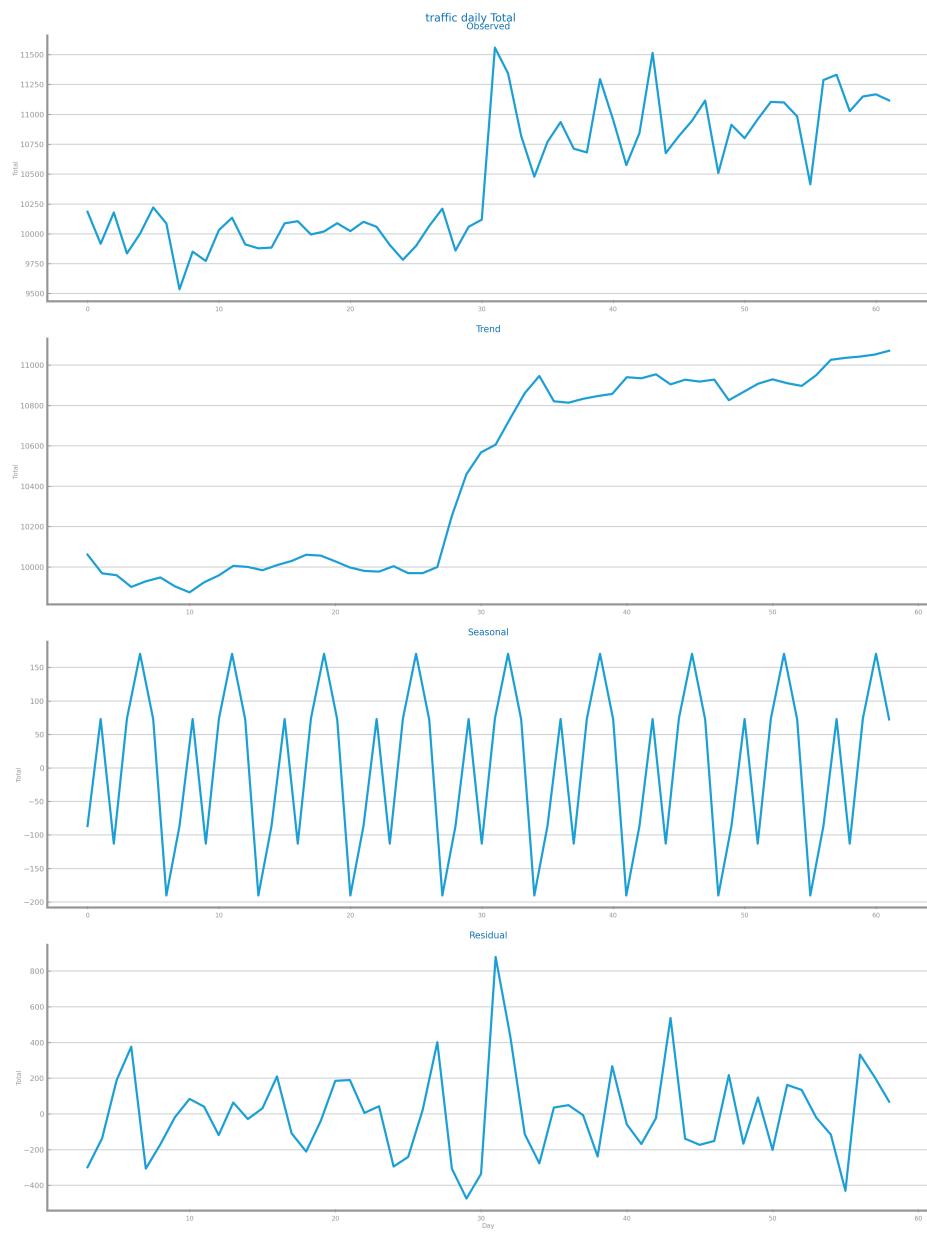


Figure 8: Components study for time series 1

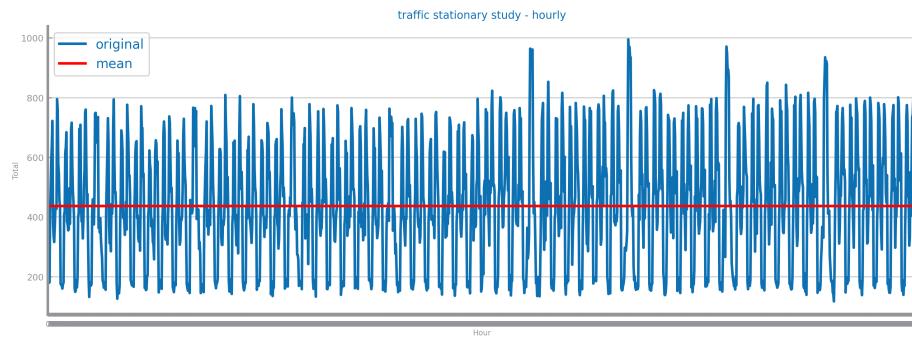


Figure 9: Stationarity study for time series 1

### **Augmented Dickey-Fuller Test Results:**

*Original (15-min):*

- ADF Statistic: -14.441
- p-value: 0.000
- Critical Values: 1%: -3.431, 5%: -2.862, 10%: -2.567
- **The series IS stationary**

*Hourly:*

- ADF Statistic: -8.903
- p-value: 0.000
- Critical Values: 1%: -3.435, 5%: -2.864, 10%: -2.568
- **The series IS stationary**

*Daily:*

- ADF Statistic: -0.826
- p-value: 0.811
- Critical Values: 1%: -3.548, 5%: -2.913, 10%: -2.594
- **The series IS NOT stationary**

## 2 PREPARATION

### 2.1 Scaling

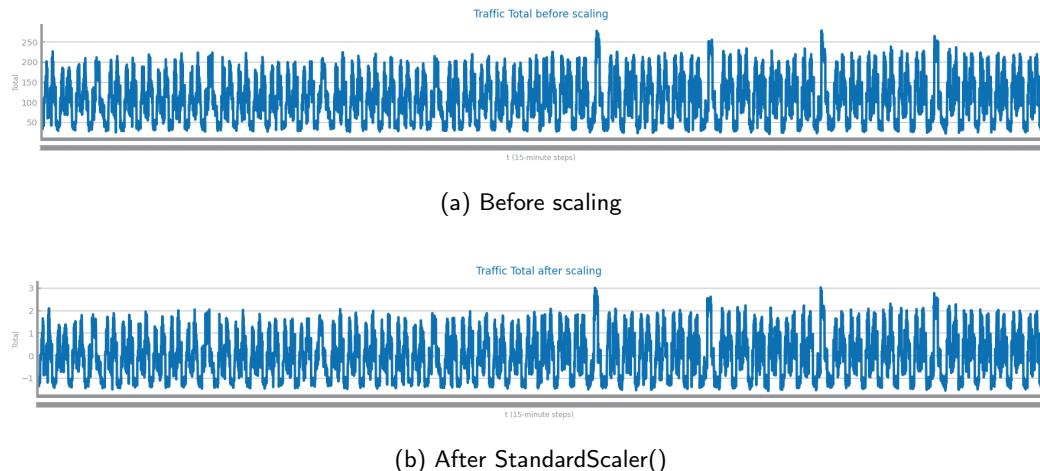
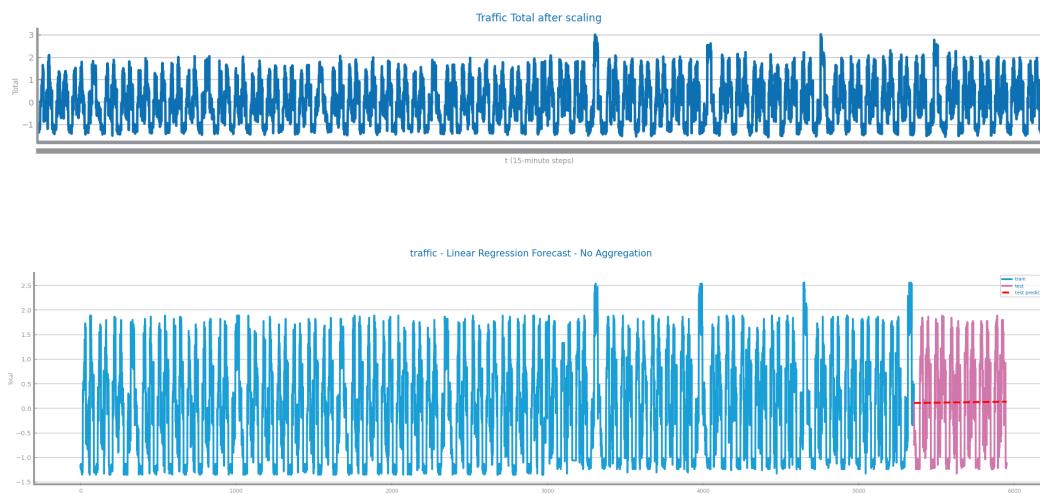


Figure 10: Effect of StandardScaler on the original 15-minute time series

### 2.2 Aggregation

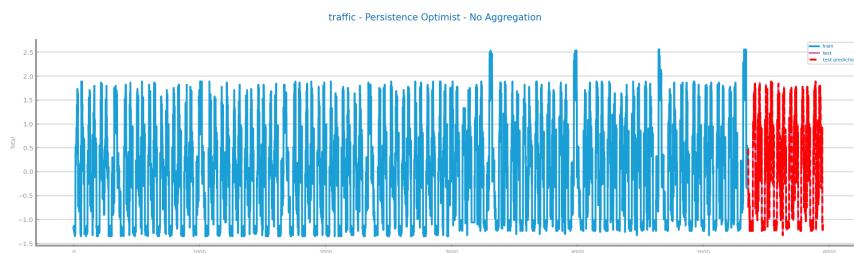
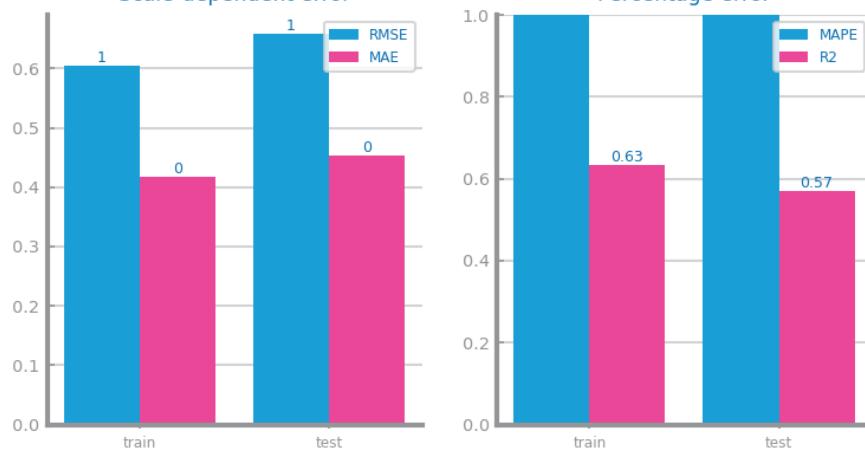
#### 2.2.1 Approach 0: No Aggregation



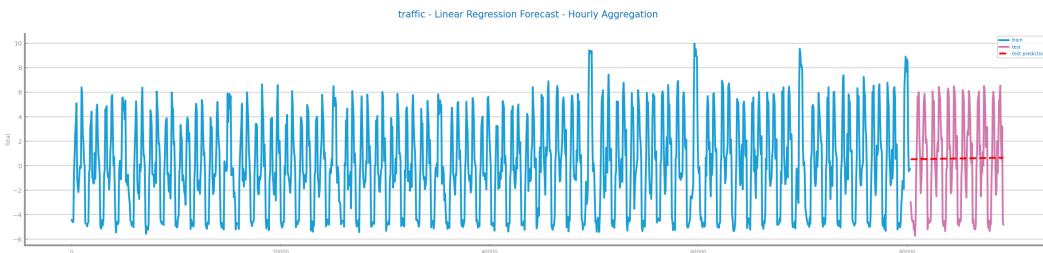
### traffic - Linear Regression Evaluation - No Aggregation



### traffic - Persistence Optimist - No Aggregation

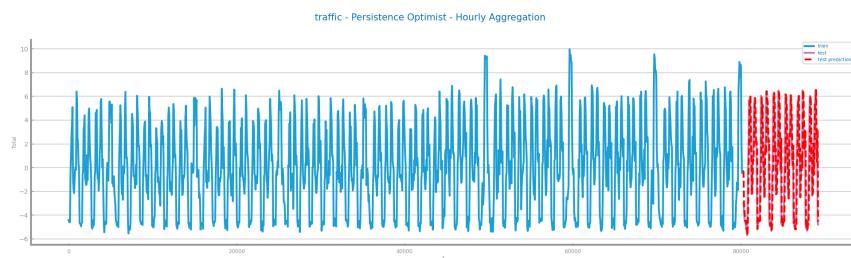
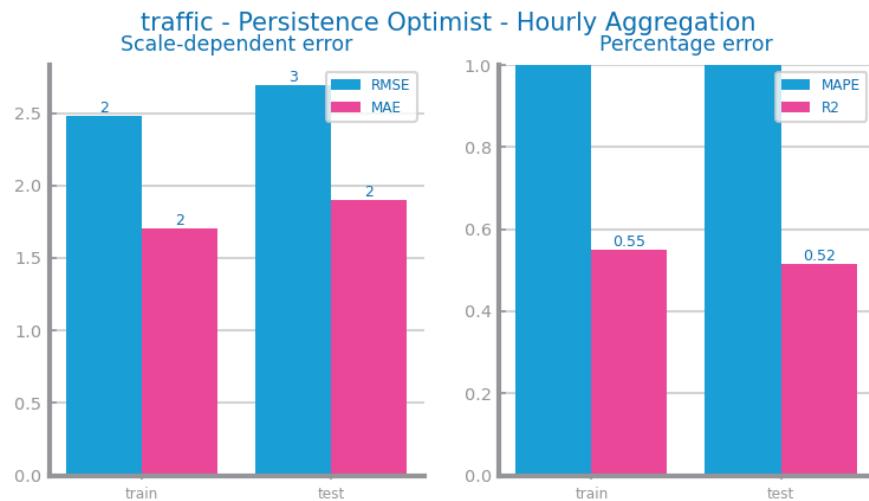


## 2.2.2 Approach 1: Hourly Aggregation

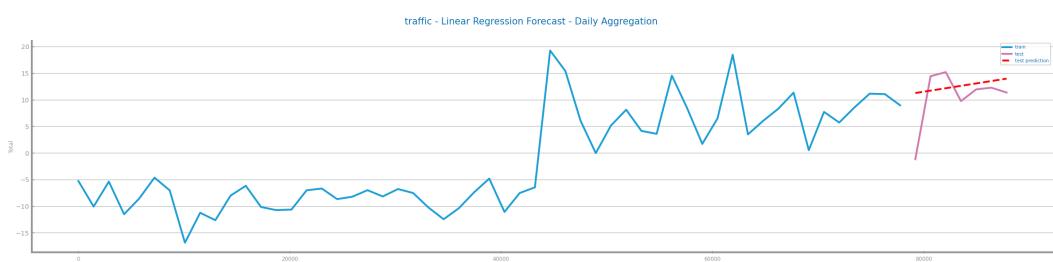


traffic - Linear Regression Evaluation - Hourly Aggregation





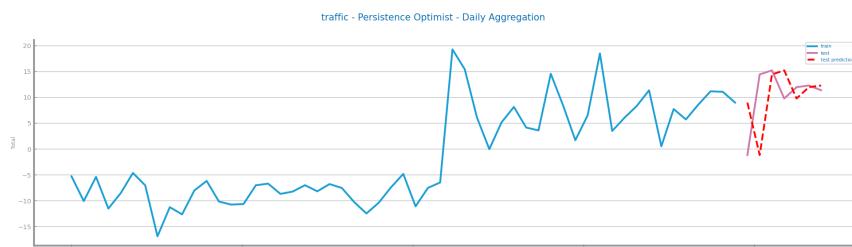
### 2.2.3 Approach 2: Daily Aggregation



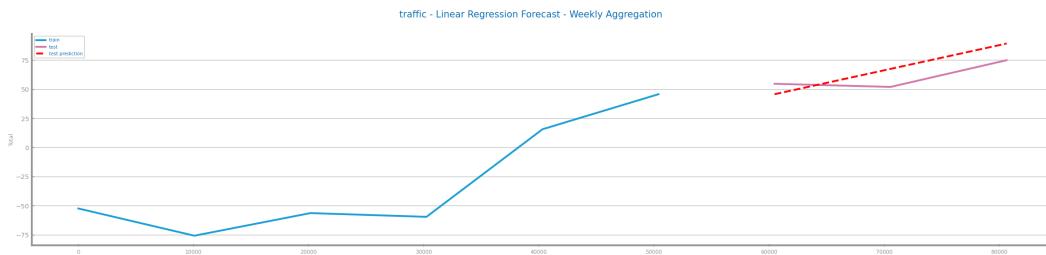
### traffic - Linear Regression Evaluation - Daily Aggregation



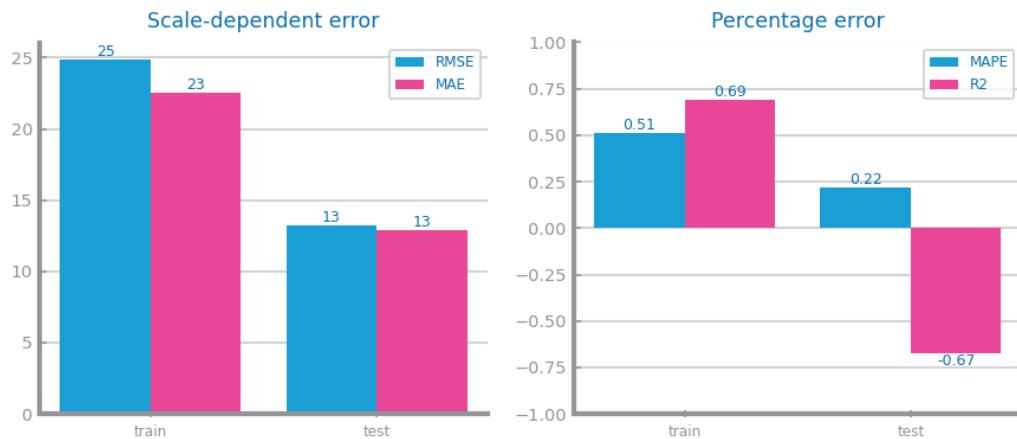
### traffic - Persistence Optimist - Daily Aggregation

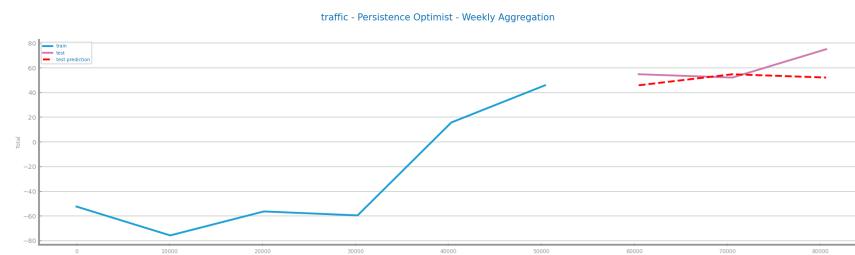
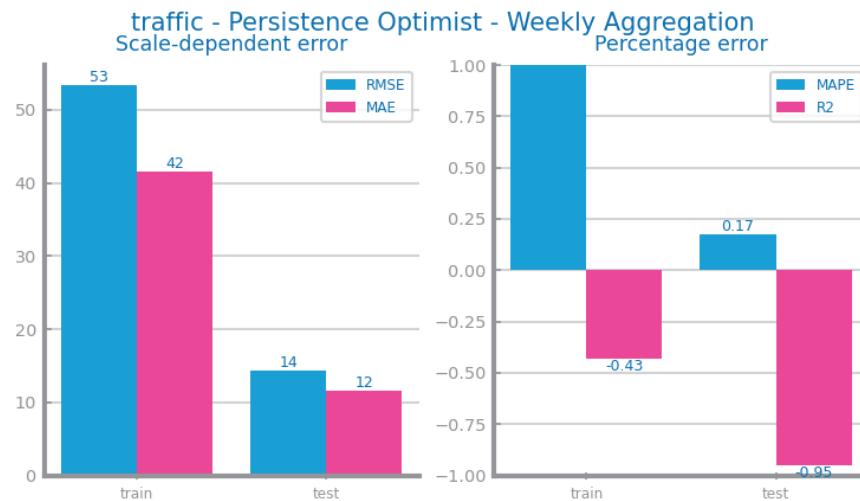


## 2.2.4 Approach 3: Weekly Aggregation

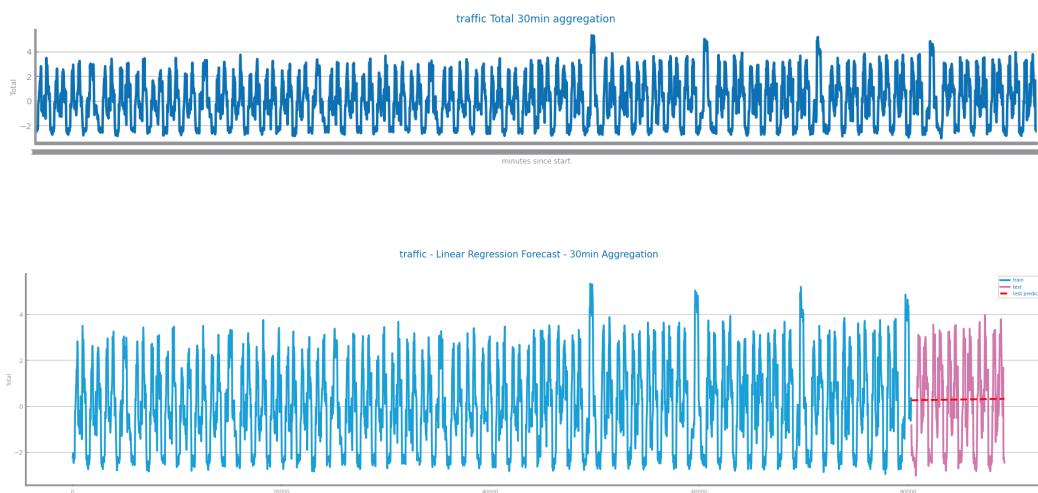


traffic - Linear Regression Evaluation - Weekly Aggregation





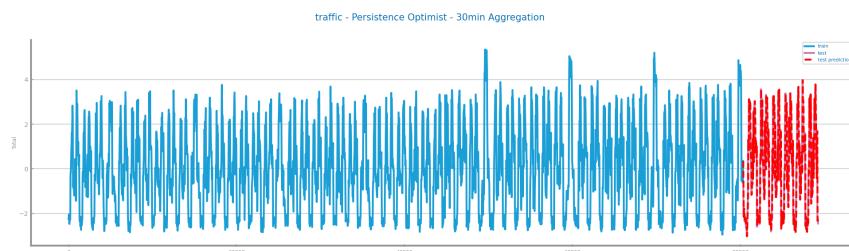
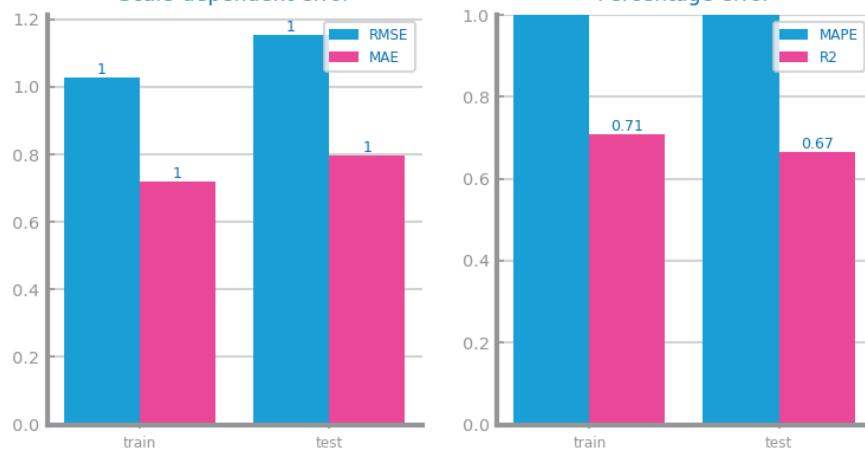
### 2.2.5 Approach 4: 30min Aggregation



### traffic - Linear Regression Evaluation - 30min Aggregation



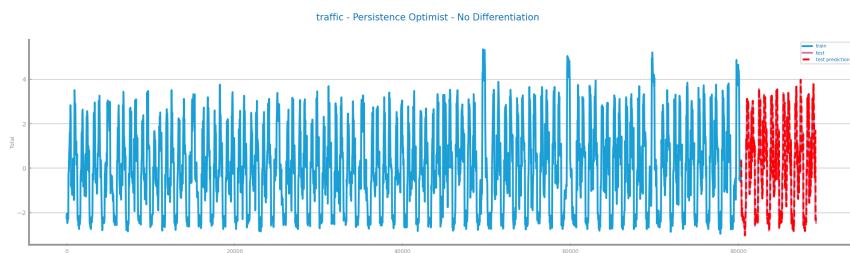
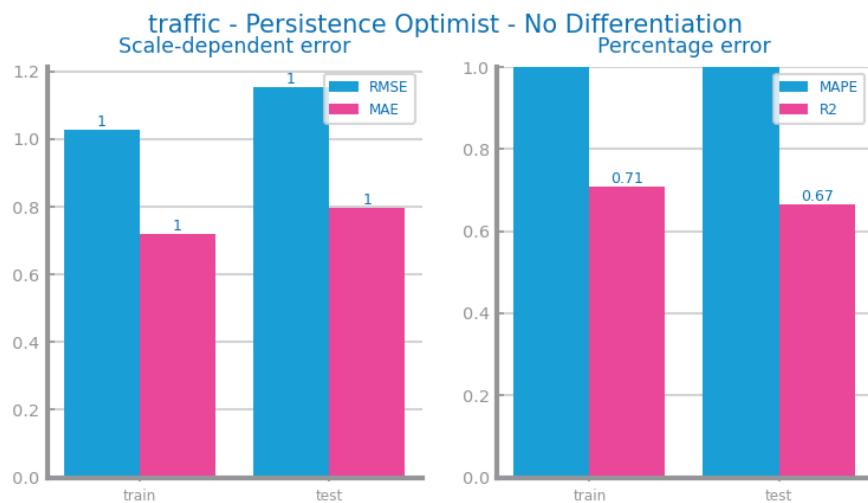
### traffic - Persistence Optimist - 30min Aggregation



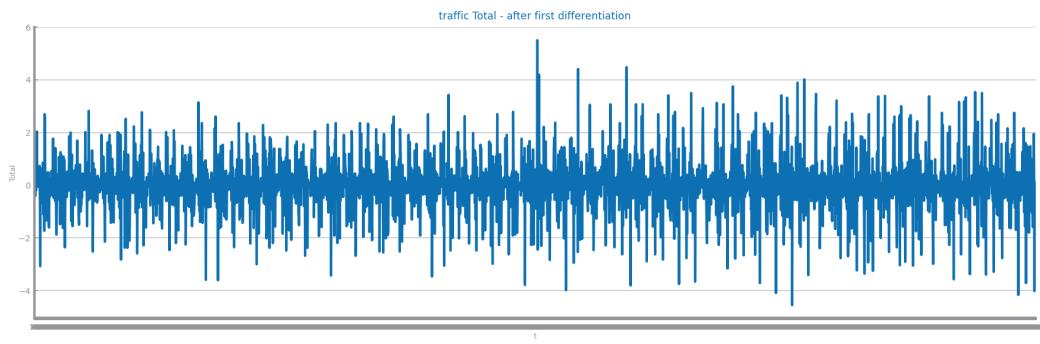
## 2.3 Differentiation

### 2.3.1 Approach 1: No differentiation





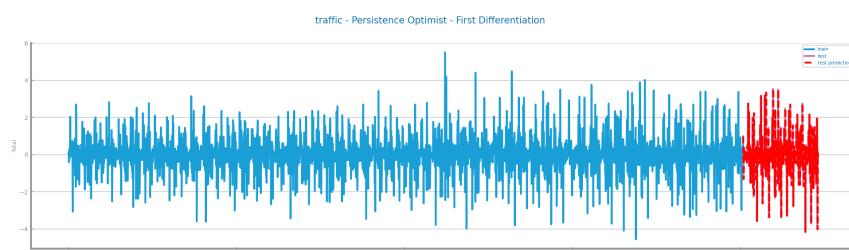
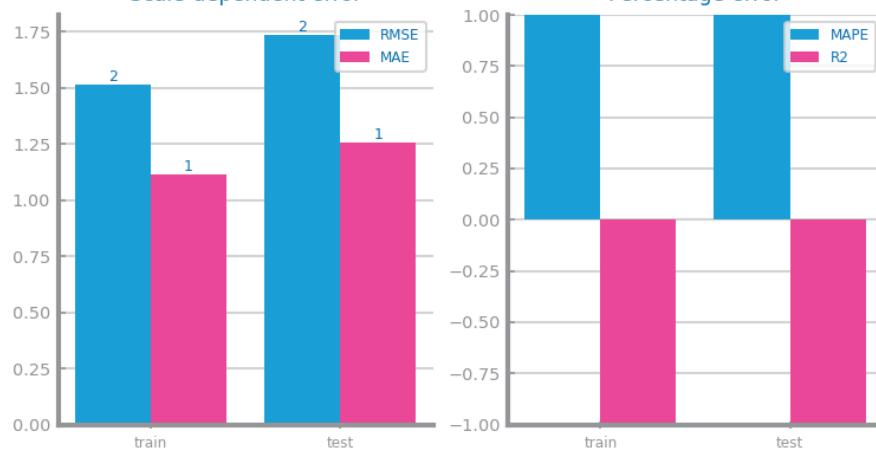
### 2.3.2 Approach 2: First differentiation



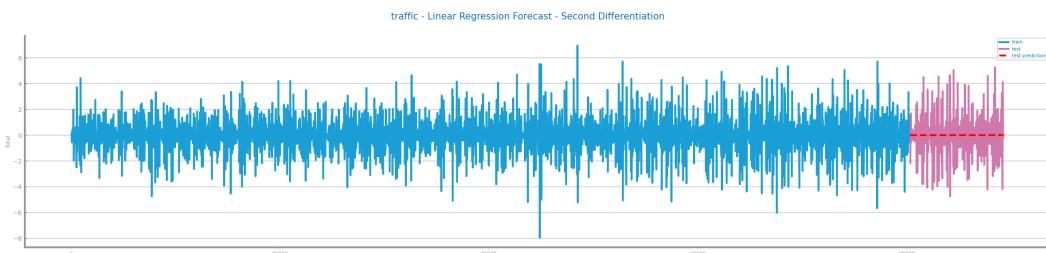
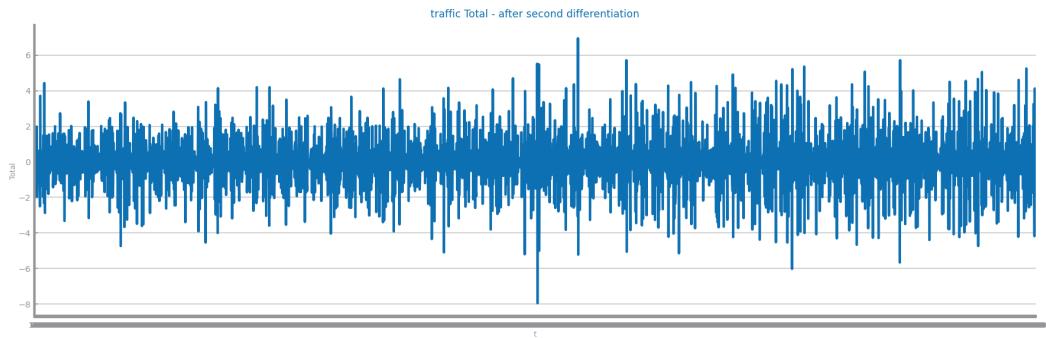
### traffic - Linear Regression Evaluation - First Differentiation



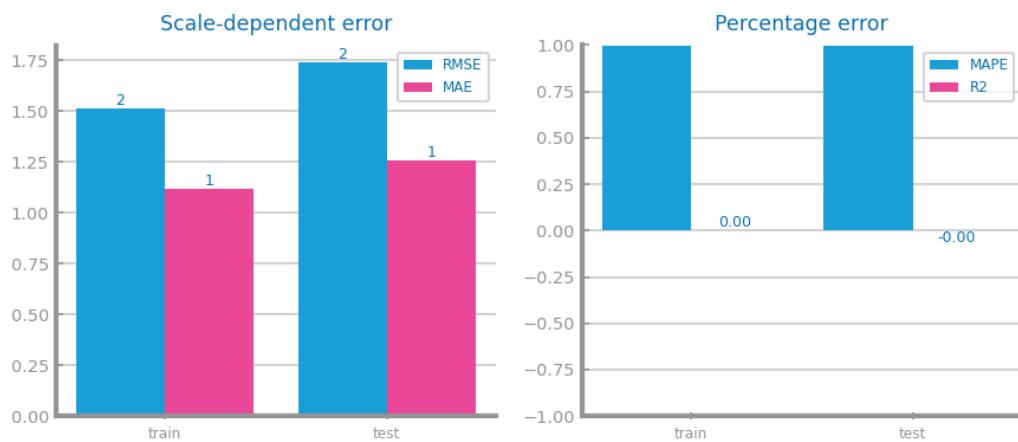
### traffic - Persistence Optimist - First Differentiation

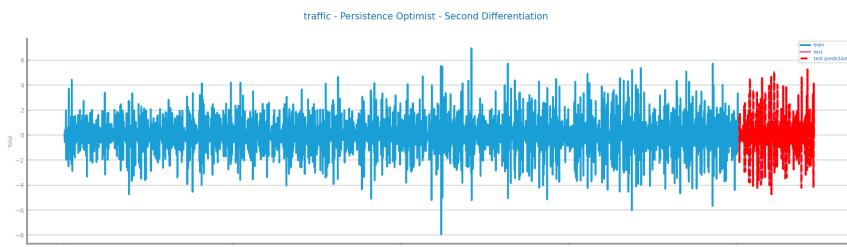


### 2.3.3 Approach 3: Second differentiation



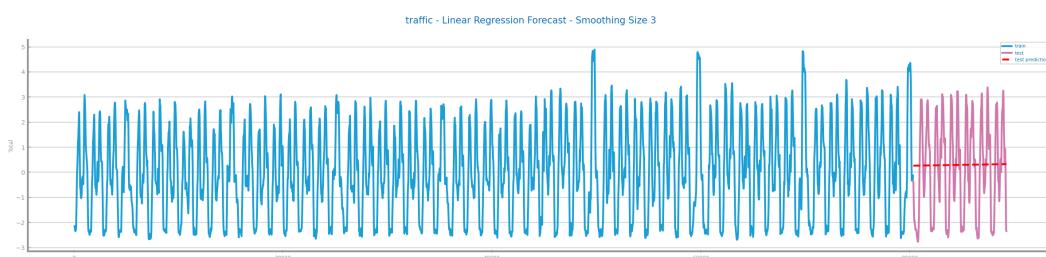
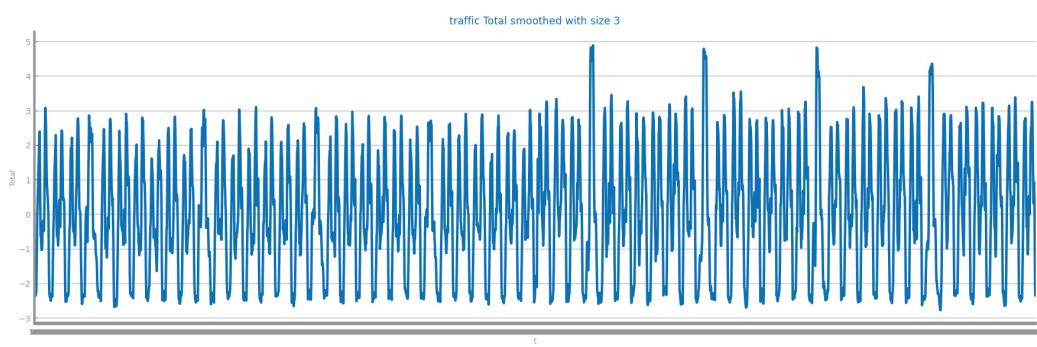
traffic - Linear Regression Evaluation - Second Differentiation





## 2.4 Smoothing

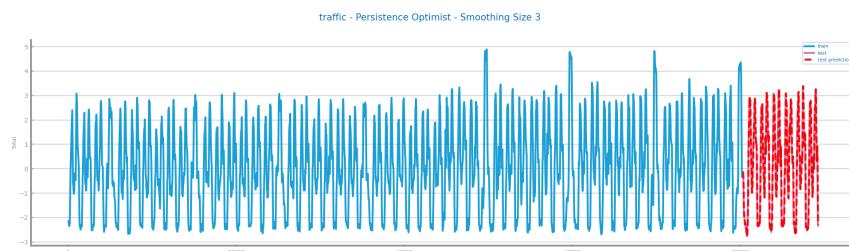
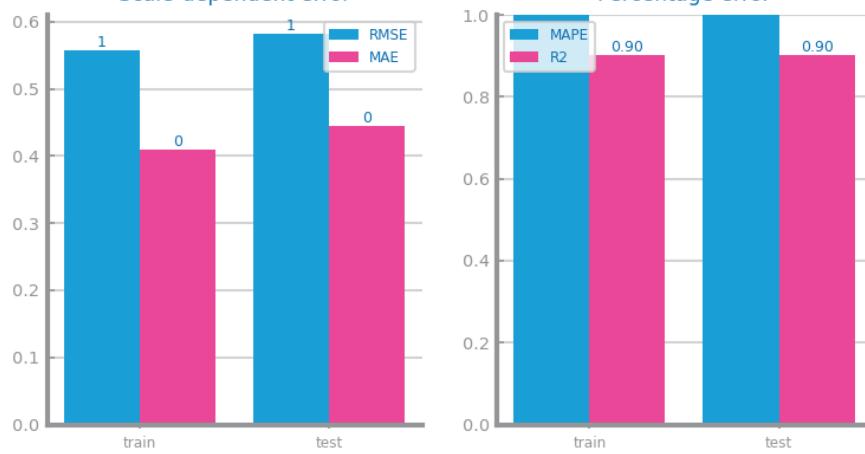
### 2.4.1 Approach 1: Smoothing Size 3



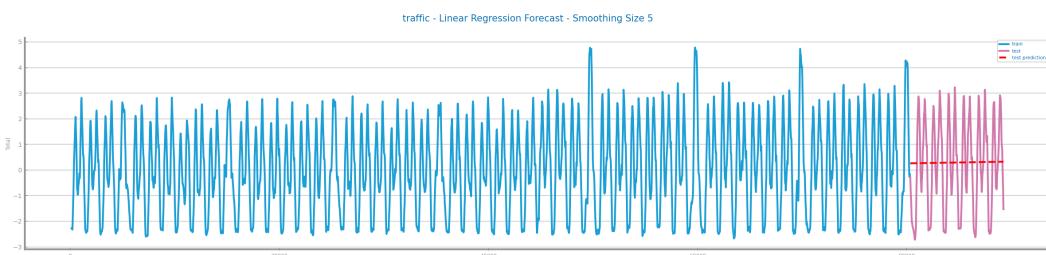
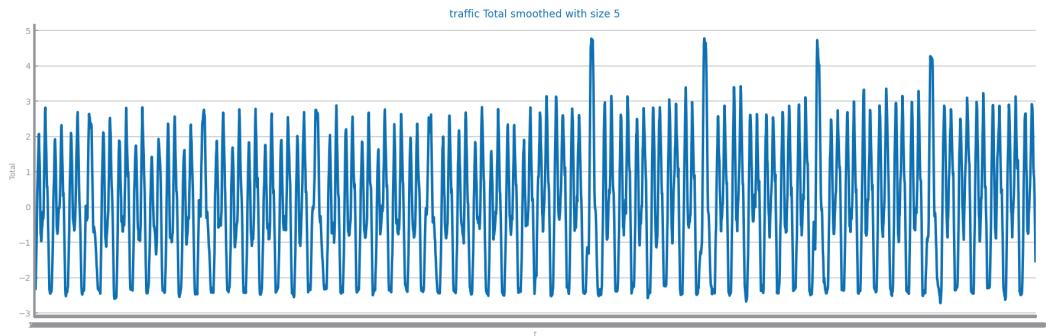
### traffic - Linear Regression Evaluation - Smoothing Size 3



### traffic - Persistence Optimist - Smoothing Size 3

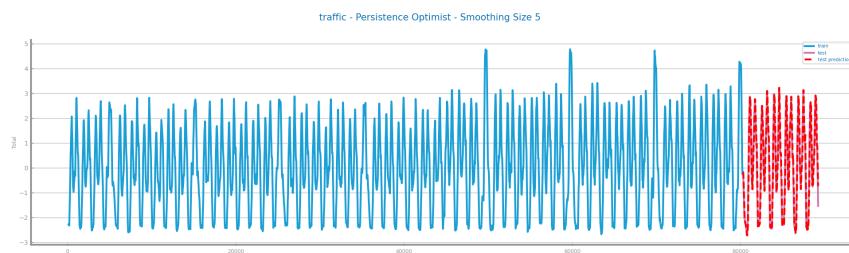


### 2.4.2 Approach 2: Smoothing Size 5

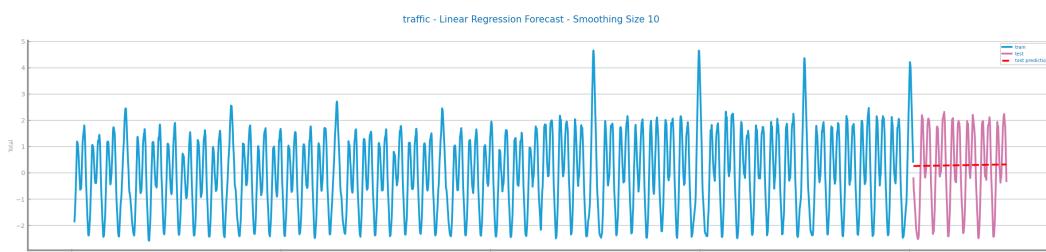
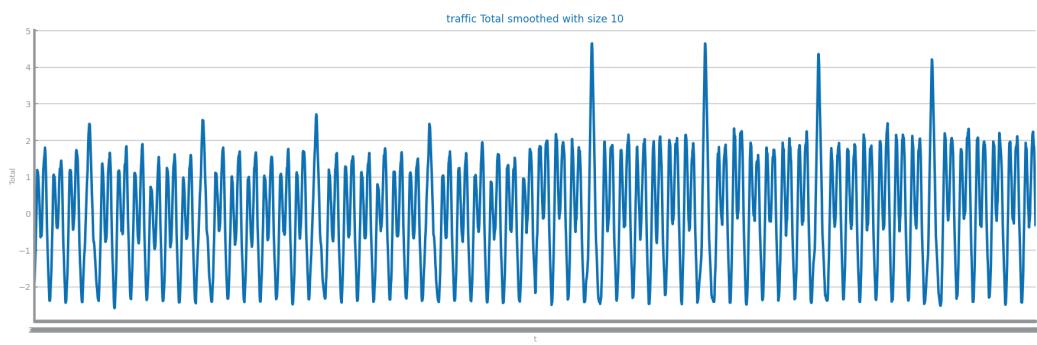


traffic - Linear Regression Evaluation - Smoothing Size 5

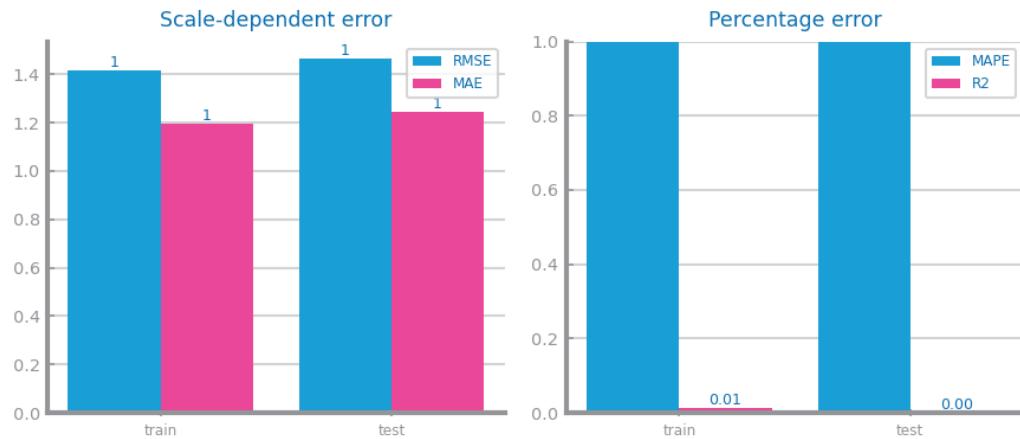




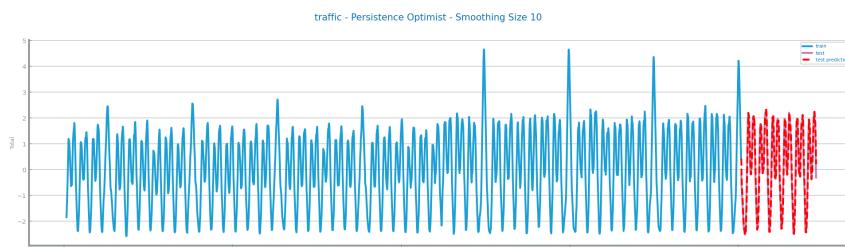
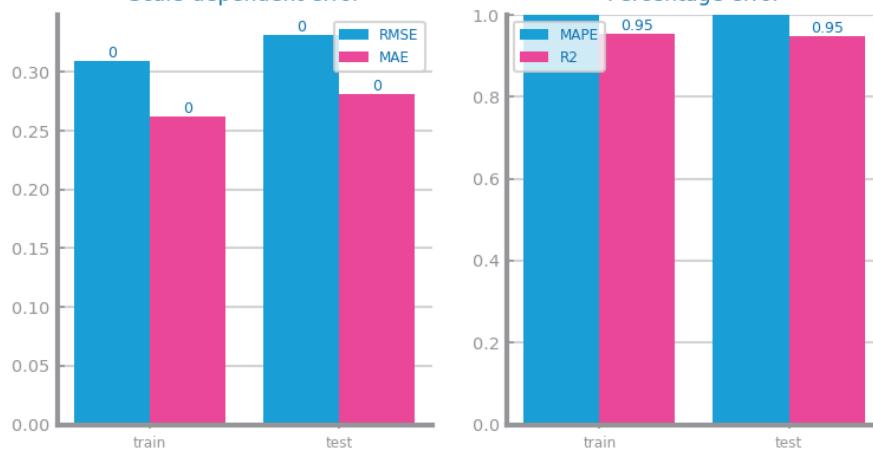
#### 2.4.3 Approach 3: Smoothing Size 10



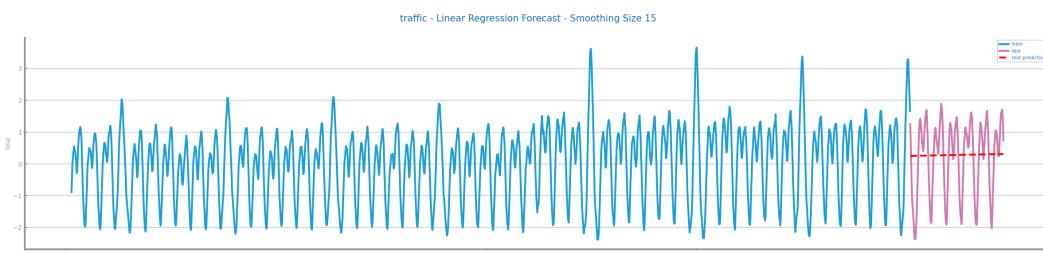
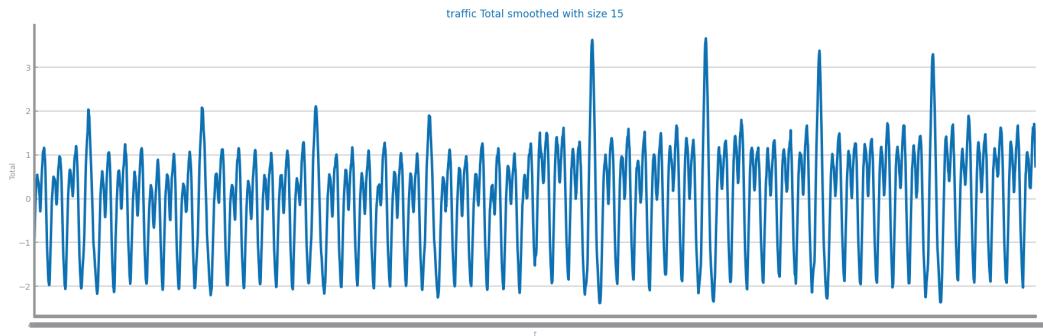
### traffic - Linear Regression Evaluation - Smoothing Size 10



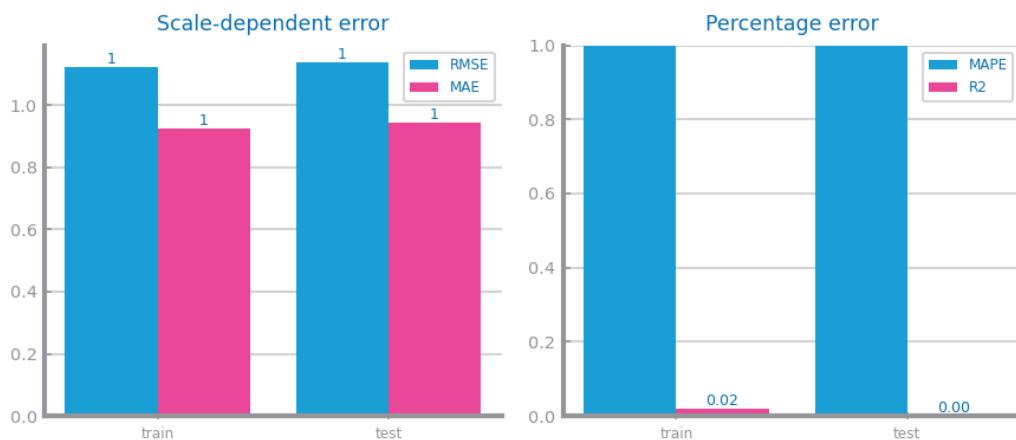
### traffic - Persistence Optimist - Smoothing Size 10

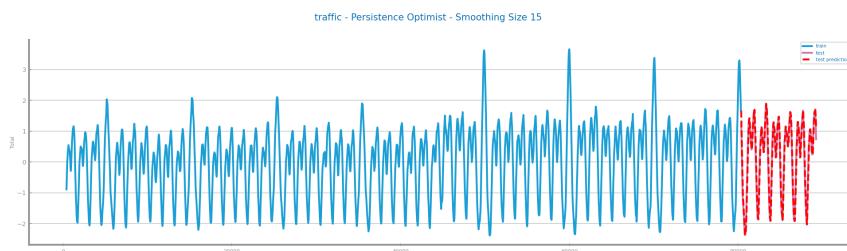


#### 2.4.4 Approach 4: Smoothing Size 15



traffic - Linear Regression Evaluation - Smoothing Size 15





## Best Configuration

Transformation	Best Configuration
Scaling	StandardScaler()
Aggregation	30-minute aggregation
Differentiation	No differentiation
Smoothing	Rolling window size = 10
<b>Best overall approach</b>	<b>30min Aggregation + No differentiation + Smoothing Size 10 + StandardScaler()</b>

### 3 Deployment

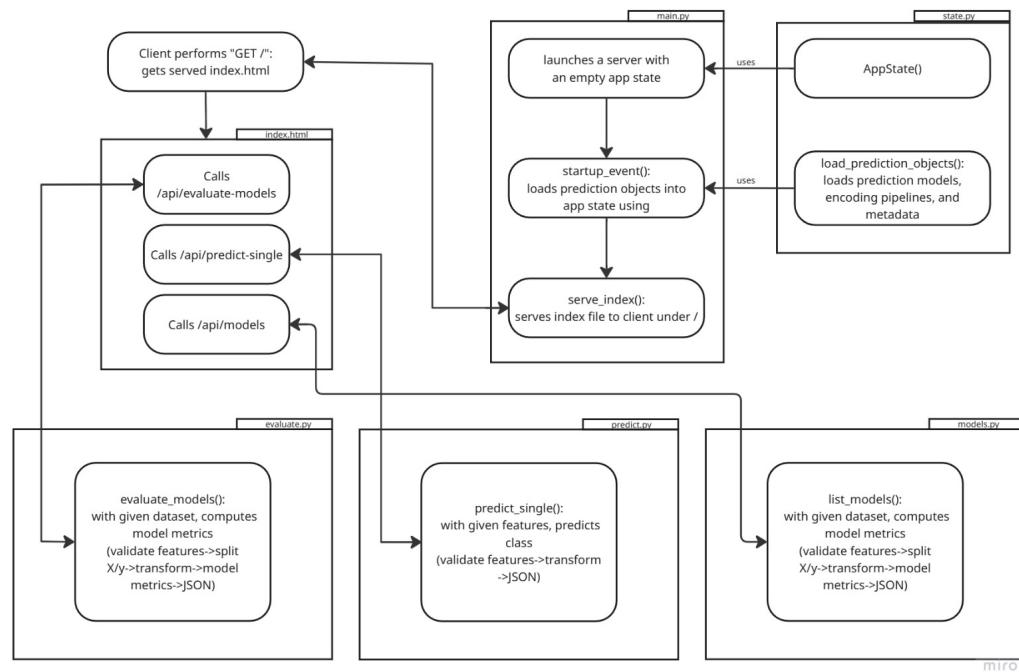


Figure 11: Architecture diagram

```
ds-deployment/
└── backend/
    ├── app/
    │   ├── main.py
    │   ├── state.py
    │   └── routes/
    │       ├── models.py
    │       ├── predict.py
    │       └── evaluate.py
    ├── models/
    │   └── traffic_lr_l1_500.joblib
    ├── pipeline/
    │   └── pipeline.joblib
    └── prediction_objects.json
└── frontend/
    └── index.html
└── requirements.txt
└── README.md
```

miro

Figure 12: File formats